

STIC-Biotech/ChemLib

92688

From: Chan, Christina
Sent: Tuesday, April 29, 2003 9:42 AM
To: Ford, Vanessa; STIC-Biotech/ChemLib
Subject: RE: Rush Sequence, In re: 09743364

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
308-3973
CM-1, 9B19

RECEIVED
APR 29 2003
(STIC)

-----Original Message-----

From: Ford, Vanessa
Sent: Monday, April 28, 2003 6:10 PM
To: Chan, Christina
Subject: Rush Sequence, In re: 09743364

Please search SEQ ID NO:1, please include interference searches.
Please rush.

Vanessa L. Ford
Biotechnology Patent Examiner
Office: CM1 8A16
Mailbox: CM1 8E12
Phone: 703.308.4735

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3504

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 4/29
Date Completed: 4/29
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 14:17:40 ; Search time 76 Seconds
(without alignments)
61.365 Million cell updates/sec

Title: US-09-743-364-1

Perfect score: 186

Sequence: 1 FTFFPPTNEIESNKKMLEKAYKESFKNGLP 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 4

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	186	100.0	35	AA176852	N-terminal fragmen
2	181	97.3	121	AAU04555	S. aureus mature c
3	58	31.2	192	ABB69541	Drosophila melanog
4	57	30.6	226	AAW20107	H. pylori cytoplas
5	57	30.6	297	AAW20698	H. pylori cytoplas
6	57	30.6	688	AA157898	Human transmembran
7	55.5	29.8	621	AA157898	Shrimp white spot
8	55	29.6	234	AA10221	Streptococcal M6'
9	55	29.6	236	AA103118	C-terminus of Stre
10	55	29.6	254	AA150228	Sequence of a frag

11	55	29.6	259	22	ABG23381	Novel human diagno
12	55	29.6	259	22	ABG23431	Novel human diagno
13	55	29.6	281	13	AA120128	Sequence encoded b
14	55	29.6	284	15	AA150229	Sequence of frame
15	55	29.6	305	15	AA150996	Recombinant M24-M5
16	55	29.6	305	21	AA103117	S. pyogenes hybrid
17	55	29.6	441	10	AA190955	M6 streptococcal p
18	55	29.6	441	14	AA141780	Streptococcus pyog
19	55	29.6	483	18	AA108927	Type-6 M-protein.
20	55	29.6	484	23	AB130015	Streptococcus poly
21	55	29.6	615	22	AA192313	C glutaminc prote
22	55	29.6	708	22	AB159644	Drosophila melanog
23	54	29.0	214	20	AA137603	Chlamydia trachoma
24	54	29.0	392	21	AA121559	Arabidopsis thalia
25	54	29.0	399	21	AA121558	Arabidopsis thalia
26	54	29.0	588	21	AA147006	Arabidopsis thalia
27	54	29.0	594	21	AA147005	Arabidopsis thalia
28	53.5	28.8	89	20	AA104699	Human RPTPa amino
29	53.5	28.8	99	20	AA104706	Mouse RPTPa amino
30	53.5	28.8	793	13	AA120743	Murine receptor ty
31	53.5	28.8	793	17	AA102282	Murine receptor ty
32	53.5	28.8	793	20	AA104695	Murine receptor ty
33	53.5	28.8	802	13	AA120744	Human receptor ty
34	53.5	28.8	802	17	AA102283	Human receptor ty
35	53.5	28.8	802	20	AA104694	Human receptor ty
36	53.5	28.8	807	23	AB142006	Human ovarian anti
37	53	28.5	386	11	AA106849	Protein Arp 4. A
38	53	28.5	386	11	AA104828	Arp 4. Streptococ
39	53	28.5	881	22	AB105280	Novel human diagno
40	53	28.5	881	22	AB120258	Novel human diagno
41	53	28.5	1847	21	AA152002	M. jannaschii MJ14
42	53	28.5	1847	21	AA151631	M. jannaschii MJ14
43	52.5	28.2	715	17	AA101461	NTHI HxuC protein.
44	52.5	28.2	1091	22	AB168898	Drosophila melanog
45	52	28.0	99	22	AA103576	Human polypeptide

ALIGNMENTS

RESULT 1
AA176852
ID AA176852 standard; peptide; 35 AA.
XX
AC AA176852;
XX
DT 12-MAY-2000 (first entry)
XX
DE N-terminal fragment of CHIPS protein.
XX
DE N-terminus; CHIPS; chemotaxis-inhibiting protein of Staphylococcus;
KW chemotaxis inhibitory protein from Staphylococcus aureus; granulocyte;
KW fMLP binding inhibitor; N-formyl-methionyl-leucyl-phenylalanine; therapy;
KW inflammation; HIV infection; Staphylococcus infection.
XX
OS Staphylococcus aureus.
XX
PN WO200002913-A1.
XX
PD 20-JAN-2000.
XX
PF 12-JUL-1999; 99WO-NL00442.
XX
PR 10-JUL-1998; 98NL-1009614.
XX
PA (EIKJ-) EIKJMAN-WINKLER INST.
XX
PI Van Strijp JAG, Van Kessel CPW;
XX
DR WPI; 2000-171131/15.
XX
PT New chemotaxis inhibiting protein of Staphylococcus CHIPS protein has
PT immunomodulating properties and is useful as general inflammation

PT inhibitor for treating AIDS and Staphylococcus infections -
 XX Claim 1; Fig 4; 28pp; English.
 XX This sequence represents the N-terminal fragment of the
 CC chemotaxis-inhibiting protein of Staphylococcus (CHIPS) of the invention.
 CC CHIPS is also referred to as chemotaxis inhibiting protein from
 CC Staphylococcus aureus. CHIPS is a 17kd protein, that is able to prevent
 CC the binding of fMLP (N-formyl-methionyl-leucyl-phenylalanine) to
 CC granulocytes. CHIPS is useful for the treatment of acute and chronic
 CC inflammation reactions and HIV infection. Antibodies against CHIPS are
 CC useful for treating Staphylococcus infection. A method for determining
 CC the chemotaxis-modulating activity of a substance can be used to identify
 CC proteins with an analogous function to CHIPS.
 XX SQ Sequence 35 AA;
 Query Match 100.0%; Score 186; DB 21; Length 35;
 Best Local Similarity 100.0%; Pred. No. 2.6e-16;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FTFFPPTNEESNKKMLEKEKAYKESFKNNGLP 35
 DQ 1 FTFFPPTNEESNKKMLEKEKAYKESFKNNGLP 35
 RESULT 2
 ID AAU04555 standard; Protein; 121 AA.
 XX AAU04555;
 DT 26-SEP-2001 (first entry)
 DE S. aureus mature chemotaxis inhibitory protein, CHIPS.
 XX Chemotaxis inhibitory protein; CHIPS; acquired immunodeficiency syndrome;
 KW AIDS; Adult Respiratory Distress syndrome; ARDS; severe infection;
 KW Crohn's disease; rheumatoid arthritis; multiple sclerosis;
 KW Alzheimer's disease.
 XX Staphylococcus aureus.
 OS WO200149711-A2.
 PN 12-JUL-2001.
 PD 08-JAN-2001; 2001WO-EP00270.
 PF 07-JAN-2000; 2000EP-0200068.
 PR (JARI-) JARI PHARM BV.
 PA Van Strijp JAG, Van Kessel CPM, Peschel AP;
 PI WPI; 2001-441844/47.
 DR N-PSDB; RAS08431.
 XX CHIPS peptides and the nucleic acids that encode them, useful for the
 PT prevention, diagnosis and treatment of e.g. human immunodeficiency
 PT virus infections and inflammation -
 XX Claim 1; Fig 5; 68pp; English.
 XX The sequence is mature S. aureus CHIPS (chemotaxis inhibitory protein).
 CC The CHIPS peptide and micro-organisms comprising the chp gene are
 CC used in prevention, diagnosis and therapy of acute and chronic
 CC inflammation reactions, human immunodeficiency virus (HIV) infections,
 CC acquired immunodeficiency syndrome (AIDS), Adult Respiratory
 CC Distress syndrome (ARDS), ischaemic shock, traumatic brain injury,
 CC severe infections, myocardial infarction, stroke, vessel surgery,
 CC ulcerative colitis, Crohn's disease, Chronic Obstructive Pulmonary
 CC disease (COPD), rheumatoid arthritis, dermatoses, multiple sclerosis,

CC Alzheimer's disease, arteriosclerosis, repetitive strain injury (RSI),
 CC acute transplant rejection, burns, acute reactive arthritis,
 CC pancreatitis, vasculitis, glomerulonephritis, gout, frost bite and/or
 CC meningitis. The CHIPS peptide may also be used in assays to identify
 CC competitors for CHIPS binding. The anti-CHIPS antibody and a CHIP
 CC receptor blocking molecule are used in the prevention, diagnosis and
 CC treatment of Staphylococcal infection. The CHIPS nucleic acids may be
 CC used in gene therapy for the treatment of inflammation and/or AIDS.
 XX SQ Sequence 121 AA;
 Query Match 97.3%; Score 181; DB 22; Length 121;
 Best Local Similarity 97.1%; Pred. No. 4.4e-15;
 Matches 34; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FTFFPPTNEESNKKMLEKEKAYKESFKNNGLP 35
 DQ 1 FTFFPPTNEESNKKMLEKEKAYKESFKNNGLP 35
 RESULT 3
 ID ABB69541 standard; Protein; 192 AA.
 XX ABB69541;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 35415.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL13644.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 35415; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 192 AA;
 Query Match 31.2%; Score 58; DB 22; Length 192;
 Best Local Similarity 52.6%; Pred. No. 17;

Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 TFEPPPTNEEIESNKKMLE 20
 ||||| |||||::: : ||
 Db 87 TFEPSPEEVDTESSLE 105

RESULT 4
 AAW20107
 ID AAW20107 standard; Protein: 226 AA.
 XX
 AC AAW20107;
 XX
 DT 08-JUL-1997 (first entry)
 XX
 DE H. pylori cytoplasmic protein, 1179838.aa.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9640893-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaard BL;
 XX
 DR WPI; 1997-052306/05.
 DR N-PSDB; AAT67356.
 XX
 XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 61; Pages 339-340; 1481pp; English.
 XX
 CC This sequence represents a H. pylori cytoplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 XX
 SQ Sequence 226 AA;

Query Match 30.6%; Score 57; DB 18; Length 226;
 Best Local Similarity 57.1%; Pred. No. 28;
 Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 7 PTNEEIESNKKMLEKAYKE 27
 :||:| || ||||| ||
 Db 136 PDPKELEQKALEKEAKE 156

RESULT 5

AAW20698
 ID AAW20698 standard; Protein: 297 AA.
 XX
 AC AAW20698;
 XX
 DT 15-JUL-1997 (first entry)
 XX
 DE H. pylori cytoplasmic protein 05ae20202orf54.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9640893-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaard BL;
 XX
 DR WPI; 1997-052306/05.
 DR N-PSDB; AAT67951.
 XX
 XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 61; Pages 1116-1117; 1481pp; English.
 XX
 CC The present sequence represents a Helicobacter pylori cytoplasmic
 CC protein which may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 XX
 SQ Sequence 297 AA;

Query Match 30.6%; Score 57; DB 18; Length 297;
 Best Local Similarity 57.1%; Pred. No. 38;
 Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 7 PTNEEIESNKKMLEKAYKE 27
 :||:| || ||||| ||
 Db 140 PDPKELEQKALEKEAKE 160

RESULT 6
 AAY57898
 ID AAY57898 standard; Protein: 688 AA.
 XX
 AC AAY57898;
 XX
 DT 23-MAR-2000 (first entry)
 XX
 DE Human transmembrane protein HTPN-22.
 XX

Human; transmembrane protein; HTPMN; diagnosis: immunospecific;
 antiproliferative; neuroprotective; immune disorder;
 reproductive disorder; smooth muscle disorder; neurological disorder;
 gastrointestinal disorder; developmental disorder;
 cell proliferative disorder.

OS Homo sapiens.

PN WO9961471-A2.

PD 02-DEC-1999.

PF 28-MAY-1999; 98WO-US11904.

PR 29-MAY-1998; 98US-0087260.

PR 02-JUL-1998; 98US-0091674.

PR 02-OCT-1998; 98US-0102954.

PR 24-NOV-1998; 98US-0109869.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;

PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR,

PI Au-Young J;

XX WPI; 2000-072605/06.

DR N-PSDB; AAZ56719.

XX Proteins, polynucleotides, vectors, host cells and antibodies used to
 PT diagnose, treat or prevent immune, reproductive, smooth muscle,
 PT neurological, gastrointestinal, developmental and cell proliferative
 PT disorders -

XX Claim 1; Page 128-130; 229pp; English.

XX AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
 CC transmembrane proteins designated HTPMN-1 to HTPMN-79, respectively.
 CC The transmembrane protein have immunospecific, antiproliferative and
 CC neuroprotective activities. The human transmembrane proteins,
 CC polynucleotides encoding them and other compositions and methods from
 CC the present invention, can be used for the diagnosis, treatment or
 CC prevention of immune, reproductive, smooth muscle, neurological,
 CC gastrointestinal, developmental and cell proliferative disorders. The
 CC HTPMN's can be used to treat or prevent disorders associated with a
 CC decreased expression or activity of HTPMN.

XX Sequence 688 AA;

Query Match 30.6%; Score 57; DB 21; Length 688;

Best Local Similarity 44.8%; Pred. No. 98;

Matches 13; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 4 EPFPTNEETESNKKMLEKAYKESFKNN 32

DB 549 EKHKKKEVEENKSKAREPPPKTKRNN 577

RESULT 7

AAG84929

ID AAG84929 standard; Protein; 621 AA.

XX AC AAG84929;

XX 11-SEP-2001 (first entry)

XX Shrimp white spot Bacilliform virus (WSBV) protein 20.

XX Shrimp white spot Bacilliform virus; WSBV; diagnosis: viral infection;
 KW antiviral agent; gene expression; antisense construct;
 KW transgenic viral resistant shrimp.

OS White spot syndrome virus.

XX

PN WO200138351-A2.

XX 31-MAY-2001.

XX 08-NOV-2000; 2000WO-US28888.

XX 24-NOV-1999; 99CN-0124717.

XX (PENY-) PE CORP NY.

PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.

PA (SINO-) SINOGENOMAX CO LTD.

XX Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;

XX WPI; 2001-355877/37.

XX N-PSDB; AAG62709.

XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus
 PT (WSBV), useful for producing viral polypeptides that can be used to
 PT screen for agents that are useful for treating WSBV infection -

XX Claim 1; Figure 3; 626pp; English.

XX The invention provides the primary nucleotide sequence of the WSBV genome
 CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
 CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
 CC molecules and proteins of the invention are useful for diagnosis and
 CC monitoring viral infection, in screens for antiviral agents and for
 CC monitoring viral gene expression or activity during a treatment regimen.
 CC The nucleic acid molecules are also useful as antisense constructs to
 CC control viral gene expression in infected cells and tissues and to create
 CC transgenic viral resistant shrimp.

XX Sequence 621 AA;

Query Match 29.8%; Score 55.5; DB 22; Length 621;

Best Local Similarity 46.2%; Pred. No. 1.3e+02;

Matches 12; Conservative 5; Mismatches 6; Indels 3; Gaps 1;

QY 10 EEIESNKKMLEKAYKESFKNGLP 35

DB 150 QDIEAEK---EKEEYEDEFKRMALP 172

RESULT 8

AAR10221

ID AAR10221 standard; Protein; 234 AA.

XX AC AAR10221;

XX 26-MAR-1991 (first entry)

XX Streptococcus M6' protein.

XX Streptococcus M protein; M'6 protein; vaccinia virus; fowlpox virus;
 KW poxviridae vaccine; streptococcal pharyngitis.

XX Streptococcus sp.

XX WO9015872-A.

XX 27-DEC-1990.

XX 21-JUN-1990; 90WO-UO03531.

XX 19-JUN-1990; 90US-0540586.

XX 21-JUN-1989; 89US-0369118.

XX (UVRO-) ROCKEFELLER UNIV.

XX Fischetti VA, Hruby DE;

XX

DR WPI; 1991-022236/03.
 XX Q-PSDB; Q10244.
 PT New recombinant streptococcal M protein DNA and viral vector -
 PT for production of poxviridae vaccines in treatment of vaccinia,
 PT fowlpox etc.
 XX Disclosure; fig 5; 41pp; English.
 XX This M'6 protein corresponds to the conserved exposed polypeptide
 CC region of the streptococcal M protein. It is encoded by a gene-
 CC tically engineered gene introduced into the genome of a vaccinia
 CC or fowlpox virus. The resultant DNA complex is useful as a vaccine
 CC for immunoprotection against streptococcal infections. The M'6
 CC polypeptide is the part of protein M responsible for virulence.
 XX Sequence 234 AA;
 SQ Query Match 29.6%; Score 55; DB 12; Length 234;
 Best Local Similarity 62.5%; Pred. NO. 52;
 Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 9 NEEIESNKKMLEKEKA 24
 DB 111 NKELESKKLLEKEKA 126
 RESULT 9
 AAB03118
 ID AAB03118 standard; Protein; 236 AA.
 XX AAB03118;
 AC AAB03118;
 XX 10-OCT-2000 (first entry)
 DT C-terminus of Streptococcus pyogenes M protein M5.
 DE Multivalent hybrid M protein; M5; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia.
 XX Streptococcus pyogenes.
 OS US6063386-A.
 XX 16-MAY-2000.
 PD 16-MAY-2000.
 XX 15-SEP-1997; 97US-0937271.
 PF 15-SEP-1997; 97US-0937271.
 XX 16-SEP-1992; 92US-0945954.
 PR 16-SEP-1992; 92US-0945954.
 XX (UYTE-) UNIV TENNESSEE RES CORP.
 PA Lederer JW, Dale JB;
 PI WPI; 2000-364475/31.
 XX New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever -
 XX Disclosure; Column 45-46; 62pp; English.
 PS The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a

CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different
 CC serotypes of group A streptococci, and may therefore be used as vaccines
 CC to protect against and control infection by type A streptococci. Type A
 CC streptococci are not only responsible for streptococcal pharyngitis
 CC (strep throat), forms of pneumonia and a condition resembling toxic
 CC shock, but are also involved in the development of acute rheumatic fever
 CC (ARF) and rheumatic heart disease. In a patient with ARF, antibodies
 CC formed during a group A streptococcal infection are also cross-reactive
 CC contain similar antigenic motifs. The new multivalent vaccines are
 CC capable of raising sero-specific antibodies against various serotypes of
 CC group A streptococci which are not cross-reactive with human heart
 CC tissue. The present sequence represents the C-terminal half of the
 CC Streptococcus pyogenes M5 protein, which is a component of the
 CC tetraivalent M24-M5-M6-M19 hybrid M protein AAB03117.
 XX Sequence 236 AA;
 SQ Query Match 29.6%; Score 55; DB 21; Length 236;
 Best Local Similarity 62.5%; Pred. NO. 52;
 Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 9 NEEIESNKKMLEKEKA 24
 DB 113 NKELESKKLLEKEKA 128
 RESULT 10
 AAR50228
 ID AAR50228 standard; Protein; 254 AA.
 XX AAR50228;
 AC AAR50228;
 XX 09-OCT-1994 (first entry)
 DT Sequence of a fragment of Group A streptococcal surface protein
 DE M5 and a carrier of the COOH-terminal portion of M5.
 XX B subunit; labile toxin; M protein; fusion protein; antigen;
 KW Group A streptococci; rheumatic fever; pharyngitis.
 XX Streptococcus Group A.
 OS WO9406465-A.
 XX 31-MAR-1994.
 PD 31-MAR-1994.
 XX 15-SEP-1993; 93WO-US08704.
 PF 15-SEP-1993; 93WO-US08704.
 XX 16-SEP-1992; 92US-0945860.
 PR 16-SEP-1992; 92US-0945860.
 XX (UYTE-) UNIV TENNESSEE RES CORP.
 PA Dale JB;
 PI WPI; 1994-118162/14.
 XX N-PSDB; AAQ45160.
 DR New recombinant hybrid streptococcal M protein antigen(s) - which
 PT elicit opsonic antibodies without eliciting cross-reactive
 PT antibodies to mammalian heart tissue
 XX Disclosure; Fig 4; 45pp; English.
 PS The surface M protein of Group A streptococci is the major virulence
 CC factor and protective antigen of these organisms. However, there are
 CC a tremendous number of M protein serotypes. The invention provides
 CC recombinant M protein antigens comprising a gene encoding a carrier
 CC protein and an NH2 or COOH terminal M protein fragment carrying one
 CC or more epitopes. The carrier may be the B subunit of E.coli labile

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (III). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 259 AA;

Query Match 29.6%; Score 55; DB 22; Length 259;

Best Local Similarity 29.5%; Pred. No. 58;

Matches 13; Conservative 7; Mismatches 14; Indels 10; Gaps 1;

QY 1 FTPEPPTNEE-----ESNKKMLEKAYKESFKNGL 34

Db 10 FAYSPMPKSNKVVKRLDDVTESYKDLLSNEADDAFAKTSSEL 53

RESULT 13

AAR20128

ID AAR20128 standard; Protein; 281 AA.

XX

AC AAR20128;

XX

DT 15-APR-1992 (first entry)

XX Sequence encoded by truncated M1 gene.

DE Protein H; immunoglobulin G; IgG; antibody; autoimmune disease.

XX Streptococcus M type 1, strain 40/58.

OS

XX Key Location/Qualifiers

FT Region 1..28

FT /label= p16M1

FT Region 29..70

FT /label= C1

FT Region 71..112

FT /label= C2

FT Region 113..155

FT /label= C3

FT Region 156..176

FT /label= C4

FT Region 177..281

FT /label= D

XX

XX WO9119740-A.

XX

XX 26-DEC-1991.

XX

XX 20-JUN-1991; 91WO-SE00447.

XX

XX 21-JUN-1990; 90SE-0002212.

XX

XX (HIGH-) HIGHTECH RECEPTOR A.

XX

XX Schmidt KH, Akesson P, Cooney J, Bjorck L;

XX

XX WPI; 1992-024366/03.

XX

XX N-PSDB; AAQ20292.

XX

XX New IgG binding proteins H' lacking an albumin binding sequence -

XX useful in purific. of excess IgG from blood and to diagnose

XX autoimmune diseases

XX

PS Disclosure; Fig 8; 37pp; English.

XX

XX The inventors claim a protein prod. by a strain of Gp A Streptococci.

XX

XX The protein has the AA sequence of protein H but lacks at least some

XX part of the C and D regions (responsible for binding albumin), esp.

XX It lacks the whole of these regions and extends for AAL to AAL58.

XX Compared with natural protein H, it is more specific and may be used

XX as part of a kit for the binding, separation and identification of

XX human IgG. The same sequences appear in WO9119741.

XX

XX Sequence 281 AA;

XX

XX

RESULT 15

AAR50996

ID AAR50996 standard; Protein; 305 AA.

Query Match 29.6%; Score 55; DB 13; Length 281;

Best Local Similarity 62.5%; Pred. No. 63;

Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 NEEIESNKKMLEKEKA 24

Db 158 NKELESKKLKEKA 173

RESULT 14

AAR50229

ID AAR50229 standard; Protein; 284 AA.

XX

AC AAR50229;

XX

XX 09-OCT-1994 (first entry)

XX

DE Sequence of fragments of Group A streptococcal surface protein

DE M5 and a carrier of the COOH-terminal portion of M5.

XX

XX B subunit; labile toxin; M protein; fusion protein; antigen;

XX Group A streptococci; rheumatic fever; pharyngitis.

XX

OS Streptococcus Group A.

XX

XX WO9406465-A.

XX

XX 31-MAR-1994.

XX

XX 15-SEP-1993; 93WO-US08704.

XX

XX 16-SEP-1992; 92US-0945860.

XX

XX (UYTE-) UNIV TENNESSEE RES CORP.

XX

XX Dale JB;

XX

XX WPI; 1994-118162/14.

XX

XX N-PSDB; AAQ45161.

XX

XX New recombinant hybrid streptococcal M protein antigen(s) - which

XX elicit opsonic antibodies without eliciting cross-reactive

XX antibodies to mammalian heart tissue

XX

XX Disclosure; Fig 5; 45pp; English.

XX

XX The surface M protein of Group A streptococci is the major virulence

XX factor and protective antigen of these organisms. However, there are

XX a tremendous number of M protein serotypes. The invention provides

XX recombinant M protein antigens comprising a gene encoding a carrier

XX protein and an NH2 or COOH terminal M protein fragment carrying one

XX or more epitopes. The carrier may be the B subunit of E. coli labile

XX toxin (LT-B) or the C-repeat portion of a streptococcal M protein.

XX

XX For example, AAQ45161/R50229 comprises three segments of M5

XX designated A, B and C. The C segment is joined by a BamHI restriction

XX site to a carrier which is the COOH-terminal half of M5. The carrier

XX includes 2.5 C-repeats, which each commence with the tetrapeptide

XX NKIS. A linker could be inserted at the BamHI site.

XX

XX Sequence 284 AA;

XX

PS Query Match 29.6%; Score 55; DB 15; Length 284;

XX

XX Best Local Similarity 62.5%; Pred. No. 64;

XX Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

XX

QY 9 NEEIESNKKMLEKEKA 24

Db 161 NKELESKKLKEKA 176

XX AAR50996;
XX AC
XX DT
XX 02-NOV-1994 (first entry)
XX DE Recombinant M24-M5-M6-M19 C-term variant.
XX KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
KW M protein; M24; M5; M6; M19; subunit; tetravalent protein; emm gene;
KW tandem; pKK223.3; protective epitope; tissue-cross-reactive epitope; A;
KW restriction enzyme site; multivalent M protein; immunisation; group A;
KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
KW antibody; heart tissue; antigen; serotype; mucosal.
XX OS
XX Synthetic.
XX PH
XX Key Location/Qualifiers
FT Peptide 1..18
FT /label= M24
FT Peptide 19..35
FT /label= M5
FT Peptide 36..52
FT /label= M6
FT Peptide 53..69
FT /label= M19
FT Peptide 70..305
FT /label= M5_C-terminal_fragment
XX W09406421-A.
XX PN
XX PD 31-MAR-1994.
XX PF 15-SEP-1993; 93WO-US08703.
XX PR 16-SEP-1992; 92US-0945954.
XX PA (UYTE-) UNIV TENNESSEE RES CORP.
XX PI Dale JB, Lederer JW;
XX WPI: 1994-118122/14.
XX DR N-PSDB; AAQ45218.
XX PT New immunogenic hybrid proteins derives from streptococcal M proteins
PT - induces opsonic antibodies, for protective immunisation against
PT against multiple group A streptococci serotypes
XX PS Disclosure; Fig 8; 67pp; English.
XX CC The sequences given in AAR50992-1001 represent hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These multivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC pKK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include
CC sequences which induce mucosal antibodies and do not require coupling
XX to an immunogenic carrier.
XX SQ Sequence 305 AA;
Query Match 29.6%; Score 55; DB 15; Length 305;
Best Local Similarity 62.5%; Pred. No. 69;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 9 NEEIESNKKMLEKKA 24
I:1:1:11:11111
DB 182 NKELESKKLTKKKA 197

Search completed: April 29, 2003, 14:26:57
Job time : 77 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 14:24:40 ; Search time 44 Seconds
(without alignments)
76.471 Million cell updates/sec

Title: US-09-743-364-1

Perfect score: 186

Sequence: 1 FTTFEPPTNEEIESNKKMLEKEKAYKESFKNNGLP 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	181	97.3	149	2	E89983
2	66.5	35.8	825	2	S54465
3	58	31.2	261	2	T09075
4	58	31.2	1650	2	T18444
5	57	30.6	431	2	T24101
6	57	30.6	445	2	T24105
7	57	30.6	522	2	H64585
8	57	30.6	522	2	B71928
9	57	30.6	1375	2	S48375
10	56.5	30.4	781	2	T02272
11	56.5	30.4	1441	2	T39636
12	56	30.1	591	2	T10640
13	55	29.6	388	2	A49545
14	55	29.6	408	2	S30283
15	55	29.6	436	2	S30284
16	55	29.6	454	2	S43556
17	55	29.6	472	2	S43554
18	55	29.6	483	2	A36297
19	55	29.6	484	2	S35401
20	55	29.6	484	2	S46489
21	55	29.6	492	2	A28616
22	55	29.6	501	2	A44643
23	55	29.6	532	2	S54871
24	55	29.6	539	2	A38549
25	55	29.6	587	2	JC1419
26	54.5	29.3	266	2	T40318
27	54	29.0	188	2	E71495
28	54	29.0	202	2	G81653
29	54	29.0	248	2	I64230

30 54 29.0 702 2 T34313
31 54 29.0 717 2 H72208
32 54 29.0 1743 2 T18279
33 53.5 28.8 168 2 G89026
34 53.5 28.8 168 2 T28776
35 53.5 28.8 181 2 A89027
36 53.5 28.8 506 2 T21941
37 53.5 28.8 668 2 T31633
38 53.5 28.8 796 1 JC1285
39 53.5 28.8 802 1 A36065
40 53.5 28.8 829 1 A47373
41 53.5 28.8 1054 2 G82934
42 53 28.5 96 2 S01250
43 53 28.5 150 2 T23684
44 53 28.5 365 2 B54128
45 53 28.5 377 2 S52537

hypothetical prote
conserved hypothet
multidrug resistan
protein T08H10.3 I
hypothetical prote
protein T08H10.4 I
hypothetical prote
hypothetical prote
protein-tyrosine-p
protein-tyrosine-p
protein-tyrosine-p
hypothetical prote
alcohol dehydrogen
hypothetical prote
Fc-binding protein
emm L 15 protein -

ALIGNMENTS

RESULT 1

E89983

hypothetical protein SAI755 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: E89983

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: E89983

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-149 <KUR>

A:Cross-references: GB:BA000018; PID:gi3701736; PIDN:BA843029.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SAI755

Query Match 97.3% Score 181; DB 2; Length 149;
Best Local Similarity 97.1%; Pred. No. 1.5e-14;
Matches 34; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 FTTFEPPTNEEIESNKKMLEKEKAYKESFKNNGLP 35

|||||

Db 29 FTTFEPPTNEEIESNKKMLEKEKAYKESFKNNGLP 63

RESULT 2

S54465

YTA12 protein precursor, mitochondrial - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YN582.14c; protein YNR089c; RCA1 protein

C:Species: Saccharomyces cerevisiae

C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 02-Feb-2001

C:Accession: S54465; S46609; A55358; S48539

R:Gentles, S.; Bowman, S.

submitted to the EMBL Data Library, May 1995

A:Reference number: S54451

A:Accession: S54465

A:Molecule type: DNA

A:Residues: 1-825 <GEN>

A:Cross-references: EMBL:Z49259; NID:g807956; PID:g807972; MIPS:YMR089c

A:Experimental source: strain AB972

R:Schwall, R.; Manhaupt, G.; Stucka, R.; Tauer, R.; Ehnlé, S.; Schwarzlose, C.; Vett

yeast 10, 1141-1155, 1994

A:Title: Identification of a set of yeast genes coding for a novel family of putative

A:Reference number: S46605; MUID:95274317; PMID:7754704

A:Accession: S46609

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-348,'EL',351-825 <SCH>
A:Cross-references: EMBL:X81068; NID:G531753; PIDN:CAA56955.1; PID:G531754
R:Tzagoloff, A.; Yue, J.; Jiang, J.; Paul, M.F.
J. Biol. Chem. 269, 26144-26151, 1994
A:Title: A new member of a family of ATPases is essential for assembly of mitochondrial
A:Reference number: A55358; MUID:95014441; PMID:7929327
A:Accession: A55358
A:Molecule type: DNA
A:Residues: 1-652,'V',654-825 <TZA>
A:Cross-references: GB:U09358; NID:G508232; PIDN:AAA62606.1; PID:G508233
C:Genetics:
A:Gene: SGD:YTA12; RCAL
A:Cross-references: SGD:S0004695; MIPS:YMR089c
A:Map position: 13R
A:Genome: nuclear
C:Superfamily: FtsH/SEC18/CDC48-type ATP-binding domain homology
C:Keywords: ATP; mitochondrion; nucleotide binding; P-loop; transmembrane protein
F:1-40/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F:41-825/Product: YTA12 protein #status predicted <MAT>
F:178-194/Domain: transmembrane #status predicted <TM1>
F:295-311/Domain: transmembrane #status predicted <TM2>
F:362-574/Domain: FtsH/SEC18/CDC48-type ATP-binding domain homology <VATP>
F:388-395/Region: nucleotide-binding motif A (P-loop)

Query Match 35.8%; Score 66.5; DB 2; Length 825;
Best Local Similarity 48.3%; Pred. No. 4.5;
Matches 14; Conservative 6; Mismatches 6; Indels 3; Gaps 1;

Oy 7 PTNEIESNKKMLEKAYKESFKNGLP 35
|||||: : ||| | | ||| :
Db 65 PTDEEVEAIRKQVER---YEQTKNTIP 90

RESULT 3
T09075
hypothetical protein CG9 (strain HB3) - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C:Accession: T09075
R:Su, X.Z.; Kirkman, L.A.; Fujioaka, H.; Wellens, T.E.
Cell 91, 593-603, 1997
A:Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resistant
A:Reference number: Z16556; MUID:98054002; PMID:9393853
A:Accession: T09075
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-261 <SUX>
A:Cross-references: EMBL:AF030690; NID:G2642510; PIDN:AAC47852.1; PID:G2642512
A:Experimental source: strain HB3; from Honduras
C:Genetics:
A:Gene: cg9

Query Match 31.2%; Score 58; DB 2; Length 261;
Best Local Similarity 37.0%; Pred. No. 14;
Matches 10; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

Oy 6 FPTNEIESNKKMLEKAYKESFKNN 32
| : || : | : || : |||
Db 83 FDEHEINNNNNKKKKSYNNYNN 109

RESULT 4
T18444
hypothetical protein C0385c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18444
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18444
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-1650 <LAW>
A:Cross-references: EMBL:Z98547; NID:el325376; PID:el427940; PIDN:CAB11112.2
C:Genetics:
A:Map position: 3
A:Introns: 1597/3; 1625/3
A:Note: C0385c

Query Match 31.2%; Score 58; DB 2; Length 1650;
Best Local Similarity 35.7%; Pred. No. 97;
Matches 10; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

Oy 5 PPTNEIESNKKMLEKAYKESFKNN 32
| : || : | : || : |||
Db 978 PVDLNRIKNETKILEKDKSHMSKIQNN 1005

RESULT 5
T24101
hypothetical protein R102.5a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24101
R:Berk, M.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19841
A:Accession: T24101
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-431 <WIL>
A:Cross-references: EMBL:Z70309; PIDN:CAA94359.1; GSPDB:GN00022; CESP:R102.5a
A:Experimental source: clone R102
C:Genetics:
A:Gene: CESP:R102.5a
A:Map position: 4
A:Introns: 21/1; 170/1; 397/3

Query Match 30.6%; Score 57; DB 2; Length 431;
Best Local Similarity 47.8%; Pred. No. 31;
Matches 11; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Oy 8 TNEIESNKKMLEKAYKESFK 30
| : ||| ||| : | : | :
Db 216 SNHEIENKKLSERMEVKNFE 238

RESULT 6
T24105
hypothetical protein R102.5b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24105
R:Berk, M.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19841
A:Accession: T24105
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-445 <WIL>
A:Cross-references: EMBL:Z70309; PIDN:CAB54290.1; GSPDB:GN00022; CESP:R102.5b
A:Experimental source: clone R102
C:Genetics:
A:Gene: CESP:R102.5b
A:Map position: 4
A:Introns: 21/1; 170/1; 397/3

Query Match 30.6%; Score 57; DB 2; Length 445;
Best Local Similarity 47.8%; Pred. No. 32;
Matches 11; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Oy 8 TNEIESNKKMLEKAYKESFK 30
| : ||| ||| : | : | :
Db 216 SNHEIENKKLSERMEVKNFE 238

```

RESULT 7
H64585
cag pathogenicity island protein cag8 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: H64585
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne-
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.;
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpik, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: H64585
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-522 <TOM>
A:Cross-references: GB:AE000567; GB:AE000511; NID:g2313641; PIDN:AA007595.1; PID:g231364
    Query Match      30.6%; Score 57; DB 2; Length 522;
    Best Local Similarity 57.1%; Pred. No. 38;
    Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 7 PTNEETESNKKMLEKAYKE 27
Db 136 PDPKELEQKKALEKEAKE 156

RESULT 8
B71928
cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: Strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: B71928
R:Alam, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: B71928
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-522 <ARN>
A:Cross-references: GB:AE001481; GB:AE001439; NID:g4155005; PIDN:AA006048.1; PID:g415500
A:Experimental source: strain J99
C:Genetics:
A:Gene: orf15
    Query Match      30.6%; Score 57; DB 2; Length 522;
    Best Local Similarity 57.1%; Pred. No. 38;
    Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 7 PTNEETESNKKMLEKAYKE 27
Db 136 PDPKELEQKKALEKEAKE 156

RESULT 9
S48375
hypothetical protein YIL159w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 06-Feb-1998
C:Accession: S48375
R:Churcher, C.
submitted to the EMBL Data Library, September 1994
A:Reference number: S48310
A:Accession: S48375
A:Molecule type: DNA
A:Residues: 1-1375 <CHU>
A:Cross-references: GB:247047; EMBL:Z38059; NID:g603997; PID:g763187; MIPS:YIL159w
C:Genetics:

```

```

A:Gene: SGD:BNR1
A:Cross-references: SGD:S0001421; MIPS:YIL159w
A:Map position: 9L
    Query Match      30.6%; Score 57; DB 2; Length 1375;
    Best Local Similarity 50.0%; Pred. No. 11e+02;
    Matches 12; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
QY 7 PTNEETESNKKMLEKAYKEKSP 30
Db 447 PINEIIOQLWKILDOKPYSES 470

RESULT 10
T02272
hypothetical protein Tl3D8.9 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 22-Oct-1999
C:Accession: T02272
R:Vysotskaia, V.S.; Schwartz, J.R.; Kwan, A.; Toriumi, M.; Yu, G.; Oji, O.; Liu, S.;
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel
submitted to the EMBL Data Library, June 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC Tl3D8 sequence.
A:Reference number: Z14649
A:Accession: T02272
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-781 <VYS>
A:Cross-references: EMBL:AC004473; NID:g3108025; PID:g3249070; GSPDB:GN00059; ATSP:T1
C:Genetics:
A:Gene: ATSP:Tl3D8.9
A:Map position: 1
A:Introns: 75/3; 126/3; 259/1; 478/1; 673/3
    Query Match      30.4%; Score 56.5; DB 2; Length 781;
    Best Local Similarity 44.8%; Pred. No. 67;
    Matches 13; Conservative 8; Mismatches 7; Indels 1; Gaps 1;
QY 8 TNEETESNKKMLEK-EKAYKESKNNGLP 35
Db 200 TDEEREADREAMEKIETAEERLKSNNPLP 228

RESULT 11
T39636
probable cleavage and polyadenylation specificity factor subunit, yeast pre-mRNA 3'-e
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39636
R:Lynne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z21868
A:Accession: T39636
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1441 <LYN>
A:Cross-references: EMBL:AL031852; PIDN:CAA21247.1; GSPDB:GN00067; SPDB:SPBC1709.08
A:Experimental source: strain 972h-; cosmid c1709
C:Genetics:
A:Gene: SPDB:SPBC1709.08
A:Map position: 2
    Query Match      30.4%; Score 56.5; DB 2; Length 1441;
    Best Local Similarity 36.7%; Pred. No. 1.3e+02;
    Matches 11; Conservative 6; Mismatches 8; Indels 5; Gaps 1;
QY 5 PFPTNEETESNKKMLEKEK-----AYKESF 29
Db 189 PYPANEDLDMEEAIAIENSISSYAYKPSF 218

RESULT 12
T10640

```

hypothetical protein T13K14.140 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
 C:Accession: T10640
 R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke
 submitted to the Protein Sequence Database, June 1999
 A:Reference number: 216991
 A:Accession: T10640
 A:Molecule type: DNA
 A:Residues: 1-591 <BEV>
 A:Cross-references: EMBL:AL080282; GSPDB:GN00062; ATSP:T13K14.140
 A:Experimental source: cultivar Columbia; BAC clone T13K14
 C:Genetics:
 A:Gene: ATSP:T13K14.140
 A:Map position: 4

Query Match 30.1%; Score 56; DB 2; Length 591;
 Best Local Similarity 39.3%; Pred. No. 57;
 Matches 11; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 5 PPTNEEIESNKKMLEKAYKESFKN 32
 | :|||:|:| :|||:|:| :||
 Db 139 PQRDEVEAKKRDAAKERARRDLN 166

RESULT 13
 A49545
 plasminogen-binding protein PAM precursor - Streptococcus pyogenes (fragment)
 N:Alternate names: plasminogen-binding M-like protein (pd 53)
 C:Species: Streptococcus pyogenes
 C>Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 26-Aug-1999
 C:Accession: A49545; S61084; S60829; S70459; S32619
 R:Berge, A.; Sjoeborg, U.
 J. Biol. Chem. 268, 25417-25424, 1993
 A:Title: PAM, a novel plasminogen-binding protein from Streptococcus pyogenes.
 A:Reference number: A49545; MUID:94064605; PMID:8244975
 A:Accession: A49545
 A:Molecule type: DNA
 A:Residues: 1-388 <BER>
 A:Cross-references: EMBL:Z22219; NID:g288978; PIDN:CAA80222.1; PID:g940870
 R:Whitmore, A.; Kapur, V.; Sullivan, D.; Musser, J.; Kehoe, M.
 submitted to the EMBL Data Library, July 1994
 A:Description: Noncongruent relationships between variation in emm gene sequences and the

A:Reference number: S61072
 A:Molecule type: DNA
 A:Residues: 13-96 <WHA>
 A:Cross-references: EMBL:U11975; NID:g533627; PIDN:AAA99591.1; PID:g12335829
 A:Experimental source: serotype M53
 R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
 Mol. Microbiol. 14, 619-631, 1994
 A:Title: Non-congruent relationships between variation in emm gene sequences and the pos
 A:Reference number: S60784; MUID:95198537; PMID:7891551
 A:Accession: S60829
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 17-77 <WHW>
 A:Cross-references: EMBL:U11975
 A:Experimental source: serotype M53
 R:Carlsson Wistedt, A.; Ringdahl, U.; Mueller-Esterl, W.; Sjoeborg, U.
 Mol. Microbiol. 18, 569-578, 1995
 A:Title: Identification of a plasminogen-binding motif in PAM, a bacterial surface prote
 A:Reference number: S70457; MUID:96342385; PMID:8748039
 A:Accession: S70459
 A:Molecule type: DNA
 A:Residues: 30-162 <CAR>
 C:Superfamily: M5 protein
 C:Keywords: cell wall
 F:1-29/Domain: signal sequence (fragment) #status predicted <SIG>
 F:30-388/Product: plasminogen-binding protein PAM #status predicted <MAT>

Query Match 29.6%; Score 55; DB 2; Length 388;
 Best Local Similarity 62.5%; Pred. No. 48;

Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 9 NEEIESNKKMLEKEKA 24
 | :|||:|:| :|||:|:| :|||
 Db 292 NAELESKALTEKEKA 307
 RESULT 14
 S30283
 protein M precursor - Streptococcus pyogenes (serotype M41)
 C:Species: Streptococcus pyogenes
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
 C:Accession: S30283; S29680
 R:Podbielski, A.
 Mol. Gen. Genet. 237, 287-300, 1993
 A:Title: Three different types of organization of the vir regulon in group A streptoc
 A:Reference number: S30283; MUID:93204905; PMID:8455563
 A:Accession: S30283
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-408 <POD1>
 A:Cross-references: EMBL:X58178
 R:Podbielski, A.; Melzer, B.
 submitted to the EMBL Data Library, February 1991
 A:Description: Pcr mediated cloning and sequencing of group A streptococcal emm41/52
 A:Reference number: S29680
 A:Accession: S29680
 A:Molecule type: DNA
 A:Residues: 1-230, 'N', 232-371, 'R', 373-408 <POD2>
 A:Cross-references: EMBL:X58178; NID:g47362; PIDN:CAA41167.1; PID:g47363
 C:Genetics:
 A:Gene: emm
 C:Superfamily: M5 protein
 C:Keywords: transmembrane protein
 F:1-41/Domain: signal sequence #status predicted <SIG>
 F:42-408/Product: M protein #status predicted <MAT>
 F:383-401/Domain: transmembrane #status predicted <TMM>
 Query Match 29.6%; Score 55; DB 2; Length 408;
 Best Local Similarity 62.5%; Pred. No. 51;
 Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 9 NEEIESNKKMLEKEKA 24
 | :|||:|:| :|||:|:| :|||
 Db 285 NKELESKALTEKEKA 300
 RESULT 15
 S30284
 M protein precursor - Streptococcus pyogenes (serotype M52)
 C:Species: Streptococcus pyogenes
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
 C:Accession: S30284; S29681
 R:Podbielski, A.
 Mol. Gen. Genet. 237, 287-300, 1993
 A:Title: Three different types of organization of the vir regulon in group A streptoc
 A:Reference number: S30283; MUID:93204905; PMID:8455563
 A:Accession: S30284
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-436 <POD1>
 A:Cross-references: EMBL:X58179
 R:Podbielski, A.; Melzer, B.
 submitted to the EMBL Data Library, February 1991
 A:Description: Pcr mediated cloning and sequencing of group A streptococcal emm41/52
 A:Reference number: S29681
 A:Accession: S29681
 A:Molecule type: DNA
 A:Residues: 1-216, 'N', 218-436 <POD2>
 A:Cross-references: EMBL:X58179; NID:g47364; PIDN:CAA41168.1; PID:g47365
 C:Genetics:
 A:Gene: emm
 C:Superfamily: M5 protein

Query Match 29.6%; Score 55; DB 2; Length 436;
Best Local Similarity 62.5%; Pred. No. 55;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 NEEIESNKKMLEKEKA 24
|:|:| :|:| | | | |
Db 313 NKEEESKKLTEKEKA 328

Search completed: April 29, 2003, 14:28:57
Job time : 45 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 14:18:05 ; Search time 27 seconds
(without alignments)

53.766 Million cell updates/sec

Title: US-09-743-364-1

Perfect score: 186

Sequence: 1 FTFFPPTNEEIESNKKMLEKAYKESFXNGLP 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	66.5	35.8	825	1 RCAL_YEAST	P40341 saccharomyc
2	57	30.6	1362	1 BRD4_HUMAN	O60885 homo sapien
3	57	30.6	1375	1 BNRI1_YEAST	P40450 saccharomyc
4	56	30.1	591	1 IF37_ARATH	P56820 arabidopsis
5	55	29.6	388	1 PM6_STRPY	P49054 streptococc
6	55	29.6	483	1 PAM_STRPY	P08089 streptococc
7	55	29.6	492	1 M5_STRPY	P02977 streptococc
8	55	29.6	539	1 M24_STRPY	P12379 streptococc
9	54	29.0	265	1 Y280_MYCGE	P47522 mycoplasma
10	54	29.0	1743	1 TAGC_DICDI	Q23868 dictyosteli
11	53.5	28.8	796	1 PTRA_RAT	Q03348 rattus norv
12	53.5	28.8	802	1 PTRA_HUMAN	P18433 homo sapien
13	53.5	28.8	829	1 PTRA_MOUSE	P18052 mus musculu
14	53	28.5	96	1 DHM2_METEX	P14775 methylobact
15	53	28.5	386	1 ARP4_STRPY	P13050 streptococc
16	53	28.5	389	1 M49_STRPY	P16947 streptococc
17	53	28.5	407	1 M21_STRPY	P50468 streptococc
18	53	28.5	804	1 MEA5_HUMAN	O15320 homo sapien
19	53	28.5	1520	1 G1TB_BACSU	P39812 bacillus su
20	52.5	28.2	741	1 BSG2_DROME	P11929 drosophila
21	52.5	28.2	791	1 Y374_TREPA	O83389 treponema p
22	52.5	28.2	886	1 RA50_SULAC	O33600 sulfolobus
23	52	28.0	564	1 M12_STRPY	P19401 streptococc
24	52	28.0	597	1 LEPA_BORPE	P56865 bordetella
25	51.5	27.7	448	1 OSH6_YEAST	Q02201 saccharomyc
26	51	27.4	230	1 YG36_YEAST	P33274 saccharomyc
27	51	27.4	451	1 TIG_HELPY	P36420 helicobacte
28	51	27.4	527	1 TGPB_YEAST	P39076 saccharomyc
29	51	27.4	596	1 RGP1_DROME	O9viw3 drosophila
30	51	27.4	1113	1 MGA2_YEAST	P40578 saccharomyc
31	51	27.4	2339	1 RFL1_PLAFA	P27625 plasmodium
32	51	27.4	6669	1 NEB9_HUMAN	P20929 homo sapien
33	50.5	27.2	199	1 MRP_MOUSE	P28667 mus musculu

RESULT 1

ID	RCAL_YEAST	STANDARD	PRT	825 AA
AC	P40341			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Mitochondrial respiratory chain complexes assembly protein RCAL (EC 3.4.24.-) (TAT-binding homolog 12).			
GN	RCAL OR YTA12 OR YMR089C OR YM9582.14C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95014441; PubMed=7929327;			
RA	Tzagoloff A., Yue J., Jang J., Paul M.F.;			
RT	"A new member of a family of ATPases is essential for assembly of-			
RT	mitochondrial respiratory chain and ATP synthetase complexes in			
RT	Saccharomyces cerevisiae."			
RL	J. Biol. Chem. 269:26144-26151(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C;			
RA	MEDLINE=95274317; PubMed=7754704;			
RA	Schnall R., Mannhaupt G., Stucka R., Tauer R., Ehnlé S.,			
RA	Schwarzlose C., Vetter I., Feldmann H.;			
RT	"Identification of a set of yeast genes coding for a novel family of			
RT	putative ATPases with high similarity to constituents of the 26S			
RT	protease complex."			
RL	Yeast 10:1141-1155(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
RA	Gentiles S., Bowman S., Barrell B.G., Rajandream M.A.;			
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.			
CC	!- FUNCTION: PUTATIVE ATP-DEPENDENT PROTEASE. NECESSARY FOR THE			
CC	ASSEMBLY OF MITOCHONDRIAL RESPIRATORY CHAIN AND ATPASE COMPLEXES.			
CC	FUNCTION BOTH IN POSTTRANSLATIONAL ASSEMBLY AND IN THE TURNOVER OF			
CC	MISTRANSLATED OR MISFOLDED POLYPEPTIDES.			
CC	!- COFACTOR: BINDS 1 ZINC ION (POTENTIAL).			
CC	!- SUBCELLULAR LOCATION: Mitochondrial.			
CC	!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.			
CC	!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; U09358; AAA62606.1; -			
DR	EMBL; X81068; CAA56955.1; -			

P15499 mus musculu
P31948 homo sapien
P30112 fasciola he
O87246 aquifex aeo
Q32356 caenorhabdi
Q9eq61 mus musculu
P40497 saccharomyc
P39936 saccharomyc
P41541 bos taurus
P40352 saccharomyc
P01211 bos taurus
P82179 canis famli

ALIGNMENTS

```

DR EMBL; 249259; CAA89236.1; -
DR MEROPS; M41.003; -
DR SGD; S0004695; YTA12.
DR InterPro; IPR003593; AAA_Atpase.
DR InterPro; IPR003959; AAA_Atpase_cent.
DR InterPro; IPR003960; AAA_sub.
DR InterPro; IPR000642; Peptidase_M41.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF01434; Peptidase_M41; 1.
DR SMART; SMO0382; AAA; 1.
DR TIGRFAMS; TIGR01241; Flsh_fam; 1.
DR PROSITE; PS00674; AAA; 1.
KW ATP-binding; Mitochondrion; Transmembrane; Hydrolase; Metalloprotease;
Zinc.
FT TRANSMEM 178 194 POTENTIAL.
FT TRANSMEM 294 311 POTENTIAL.
FT NP_BIND 388 395 ATP (POTENTIAL).
FT METAL 613 613 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 614 617 ZINC (CATALYTIC).
FT METAL 617 617 ZINC (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 349 350 DV -> EL (IN REF. 2).
FT CONFLICT 653 653 I -> V (IN REF. 1).
SQ SEQUENCE 825 AA; 93276 MW; 63CEB9EF11B3DFC CRC64;

Query Match 35.8%; Score 66.5; DB 1; Length 825;
Best Local Similarity 48.3%; Pred. No. 1.8;
Matches 14; Conservative 6; Mismatches 6; Indels 3; Gaps 1;

Oy 7 PTNEIESNKKMLEKEKAYKESFKNNGLP 35
Db 65 PTDEEVEAIRKOVER---YIEQTKNTIP 90

RESULT 2
BRD4_HUMAN
ID BRD4_HUMAN STANDARD; PRT; 1362 AA.
AC O60885; Q36PD3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bromodomain-containing protein 4 (HUNK1 protein).
GN BRD4 OR HUNK1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA French C.A., Fletcher J.A.;
RT "Human BRD4 protein.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-722 FROM N.A.
RA Weber B.;
RT TISSUE=Placenta;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF386649; AAL26987.1; -
CC EMBL; Y12059; CAA72780.1; -
CC HSSP; Q92831; 1B91.
CC Genew; HGNC:13575; BRD4.
CC InterPro; IPR001487; Bromodomain.
CC Pfam; PF00439; Bromodomain; 2.

```

```

DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SMO0297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00104; BROMODOMAIN_2; 2.
KW Bromodomain; Repeat; Nuclear protein.
FT DOMAIN 75 147 BROMODOMAIN 1.
FT DOMAIN 368 440 BROMODOMAIN 2.
FT DOMAIN 535 594 LYS-RICH.
FT DOMAIN 692 717 SER-RICH.
FT DOMAIN 703 714 POLY-SER.
FT DOMAIN 738 743 POLY-HIS.
FT DOMAIN 757 761 POLY-PRO.
FT DOMAIN 764 770 POLY-PRO.
FT DOMAIN 771 775 POLY-GLN.
FT DOMAIN 776 783 POLY-PRO.
FT DOMAIN 954 964 POLY-PRO.
FT DOMAIN 974 986 POLY-PRO.
FT DOMAIN 1011 1014 POLY-PRO.
FT DOMAIN 1028 1033 POLY-PRO.
FT DOMAIN 1283 1300 POLY-GLN.
FT DOMAIN 1301 1308 POLY-ALA.
FT DOMAIN 1335 1338 POLY-ARG.
FT CONFLICT 720 721 EM -> GP (IN REF. 2).
SQ SEQUENCE 1362 AA; 152219 MW; D52EFEICF9960907 CRC64;

Query Match 30.6%; Score 57; DB 1; Length 1362;
Best Local Similarity 44.8%; Pred. No. 42;
Matches 13; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

Oy 4 EFFTNEIESNKKMLEKEKAYKESFKNN 32
Db 549 EKHKRKEEVEENKSKAKEPPPKTKKNN 577

RESULT 3
BNRI_YEAST
ID BNRI_YEAST STANDARD; PRT; 1375 AA.
AC P40450;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE BNI1 related protein 1.
GN BNI1 OR YIL159W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Holsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moutle T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP FUNCTION.
RX MEDLINE=97327553; PubMed=9184220;
RA Inamura H., Tanaka K., Hihara T., Umikawa M., Kamei T., Takahashi K.,
RA Sasaki T., Takai Y.;
RT "Bni1p and Bnr1p: downstream targets of the Rho family small
RT G-proteins which interact with profilin and regulate actin
RT cytoskeleton in Saccharomyces cerevisiae.";
RL EMBL J. 16:2745-2755(1997).
CC -!- FUNCTION: MAY ORGANIZE MICROTUBULES BY MEDIATING SPINDLE
CC POSITIONING AND MOVEMENT IN THE BUDDING PROCESS. POTENTIAL TARGET
CC OF THE RHO FAMILY MEMBERS (BY SIMILARITY).
CC -!- SUBUNIT: INTERACTS WITH PROFILIN AT THE FHI DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 1 (FHI) DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 2 (FH2) DOMAIN.

```

```

CC  -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.
CC  -!- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).
CC  -!- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. BN1L
CC  SUBFAMILY.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Z38059; CAA86119.1; -
DR  PIR; S48375; S48375.
DR  SGD; S0001421; BN1L.
DR  InterPro; IPR003104; FH2.
DR  Pfam; PF02181; FH2; 1.
DR  SMART; SM00498; FH2; 1.
KW  Coiled coil.
FT  DOMAIN 82 284
FT  PDB 3B3.
FT  DOMAIN 180 513
FT  COILED COIL (POTENTIAL).
FT  DOMAIN 520 601
FT  FH1 (PRO-RICH).
FT  DOMAIN 659 851
FT  PRO-RICH.
FT  DOMAIN 839 839
FT  FH2.
FT  DOMAIN 868 1332
FT  DAD.
FT  DOMAIN 1316 1330
FT  ARG/LYS-RICH (BASIC).
SQ  SEQUENCE 1375 AA; 156851 MW; B16B04A33C9D2F08 CRC64;

Query Match 30.6%; Score 57; DB 1; Length 1375;
Best Local Similarity 50.0%; Pred. NO. 42;
Matches 12; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 7 PTNEIESNKKMLEKAYKESFK 30
Db 447 PINEIQLSKWILDSQKPYSESIK 470

RESULT 4
ID IF37_ARATH STANDARD; PRT; 591 AA.
AC P56620;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative eukaryotic translation initiation factor 3 subunit 7 (eIF-3
DE zeta).
GN AtG20980 OR T13K14.140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidtheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hebeisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.A., McCullagh B., Billam L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Bracken M., Waltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenecker T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moolijm P., Klein lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berneriser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buysschaert C., Gielen J., Villarroel R., De Clercq R.,

```

```

RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloecker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Farmann B., Graenderath K., Dauner D., Herzi A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Massenat O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Cheford F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Baryes M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Thuideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antonoli B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana";
RL Nature 402:769-777(1999).
CC -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA. ASSOCIATES WITH THE SUBUNIT P170 OF
CC EIF3 (BY SIMILARITY).
CC -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EIF37 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AL080282; CAB45893.1; -
DR  EMBL; AL161534; CAB79098.1; -
KW  Hypothetical protein; Initiation factor; Protein biosynthesis.
SQ  SEQUENCE 591 AA; 66724 MW; 9C5E673F04E9845C CRC64;

Query Match 30.1%; Score 56; DB 1; Length 591;
Best Local Similarity 39.3%; Pred. NO. 22;
Matches 11; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 5 PFPTNEIESNKKMLEKAYKESFKN 32
Db 139 PORRDEEVEAKKRAEKERARDRLYN 166

RESULT 5
ID PAM_STRPY STANDARD; PRT; 388 AA.
AC P49054;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Plasminogen-binding protein pam precursor (Fragment).
GN PAM.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;

```

```

RN SEQUENCE FROM N.A., AND SEQUENCE OF 32-41.
RC STRAIN-AP53 / Serotype M53;
RX MEDLINE=94064605; PubMed=8244975;
RA Berge A., Sjoebing U.;
RT "PAM, a novel plasmidogen-binding protein from Streptococcus
  pyogenes.";
RL J. Biol. Chem. 268:25417-25424(1993).
CC -!- FUNCTION: BINDS TO HUMAN PLASMINOGEN. COULD PROVIDE THE BACTERIA
  WITH A MECHANISM FOR INVASION.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
  an amide bond (potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 222219; CAA80222.1;
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF02370; M; 8.
KW Cell wall; Peptidoglycan-anchor; Repeat; Signal.
FT NON_TER 1
FT SIGNAL <1 29
FT CHAIN 30 384
FT PROPEP 385 >388
FT DOMAIN 91 116
FT REPEAT 91 103
FT REPEAT 104 116
FT DOMAIN 147 161
FT DOMAIN 147 153
FT REPEAT 147 153
FT REPEAT 154 161
FT DOMAIN 163 278
FT REPEAT 163 204
FT REPEAT 205 246
FT REPEAT 247 278
FT DOMAIN 344 380
FT SITE 381 385
FT MOD_RES 384 384
FT NON_TER 388 388
SQ SEQUENCE 388 AA; 43629 MW; EEEC4FD962CCDB12 CRC64;

Query Match 29.6%; Score 55; DB 1; Length 388;
Best Local Similarity 62.5%; Pred. No. 19;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 NEEIESNKKMLEKEKA 24
DB 292 NKELESKKLLEKEKA 307

RESULT 6
M6_STRPY
ID M6_STRPY STANDARD; PRT; 483 AA.
AC P09089;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE M protein, serotype 6 precursor.
GN EMW6.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86111835; PubMed=3511046;
RA Hollingshead S.K., Fischetti V.F., Scott J.R.;

```

```

RT "Complete nucleotide sequence of type 6 M protein of the group A
  Streptococcus. Repetitive structure and membrane anchor.";
RL J. Biol. Chem. 261:1677-1686(1986).
RN [2]
RP SEQUENCE OF 43-122 FROM N.A.
RX MEDLINE=85166224; PubMed=3885219;
RA Scott J.R., Pulliam W.M., Hollingshead S.K., Fischetti V.A.;
RT "Relationship of M protein genes in group A streptococci.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:1822-1826(1985)
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
  OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
  THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
  PHAGOCYTOSIS.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
  an amide bond (potential).
CC -!- SIMILARITY: TO OTHER M PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11338; AAA26920.1;
DR PIR; A26297; A26297.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02370; M; 9.
DR PRINTS; PR00015; GPOSANCHOR.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs; TIGR01168; VSIRK_signal; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Virulence; Phagocytosis; Cell wall; Peptidoglycan-anchor; Repeat;
  Antigen; Coiled coil; Signal.
FT SIGNAL 1 42
FT CHAIN 43 452
FT PROPEP 453 483
FT DOMAIN 69 138
FT DOMAIN 157 269
FT DOMAIN 279 347
FT DOMAIN 348 411
FT DOMAIN 412 448
FT SITE 449 453
FT MOD_RES 452 452
FT SEQUENCE 483 AA; 53472 MW; 68F87F2BDB53A448 CRC64;

Query Match 29.6%; Score 55; DB 1; Length 483;
Best Local Similarity 62.5%; Pred. No. 24;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 NEEIESNKKMLEKEKA 24
DB 360 NKELESKKLLEKEKA 375

RESULT 7
M5_STRP5
ID M5_STRP5 STANDARD; PRT; 492 AA.
AC P02977;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE M protein, serotype 5 precursor.
GN EMW5 OR SMP5.
OS Streptococcus pyogenes (serotype M5).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=160491;
RN [1]

```


Db 416 NKELESKMLTEKEKA 431

RESULT 9

Y280_MYCGE STANDARD; PRT; 265 AA.

AC P47522;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein MG280.

GN MG280.

OS Mycoplasma genitalium.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=APCC 33530 / G-37;

RX MEDLINE=96026346; PubMed=756993;

RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,

RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,

RA Eickbush J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,

RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,

RA Tomb J.-F., Dougherty B.A., Boff K.F., Hu P.-C., Lucier T.S.,

RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;

RT "The minimal gene complement of Mycoplasma genitalium";

RL Science 270:397-403(1995).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U60086; AAB03331.1;

DR DictyDB; D02060; tagc.

DR InterPro; IPR003593; AAA_AtpPase.

DR InterPro; IPR003439; ABC_transportr.

DR InterPro; IPR001140; ABCtransportr.

DR InterPro; IPR000209; Peptidase_S8.

DR Pfam; PF00005; ABC_tran; 1.

DR Pfam; PF00664; ABC_membrane; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR SMART; SM00382; AAA; 1.

DR PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.

DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; FALSE_NEG.

DR PROSITE; PS00211; ABC_TRANSPORTER; 1.

KW Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;

KW Signal.

FT SIGNAL 1 27

FT CHAIN 28 1743

FT DOMAIN 316 642

FT DOMAIN ? 1743

FT TRANSMEM 962 982

FT TRANSMEM 1027 1047

FT TRANSMEM 1072 1092

FT TRANSMEM 1157 1177

FT TRANSMEM 1260 1280

FT TRANSMEM 1288 1308

FT ACT_SITE 325 325

FT ACT_SITE 372 372

FT ACT_SITE 637 637

FT NP_BIND 1495 1492

FT DOMAIN 42 46

FT DOMAIN 94 103

FT DOMAIN 643 646

FT DOMAIN 733 741

FT DOMAIN 786 792

FT DOMAIN 1337 1340

FT DOMAIN 1346 1352

FT DOMAIN 1353 1357

FT DOMAIN 1358 1364

FT DOMAIN 1381 1386

FT DOMAIN 1707 1729

FT CARBOHYD 390 390

FT CARBOHYD 536 536

FT CARBOHYD 547 547

FT CARBOHYD 614 614

FT CARBOHYD 689 689

FT CARBOHYD 735 735

FT CARBOHYD 741 741

FT CARBOHYD 776 776

FT CARBOHYD 832 832

FT CARBOHYD 887 887

FT CARBOHYD 1251 1251

FT CARBOHYD 1385 1385

FT CARBOHYD 1386 1386

FT CARBOHYD 1454 1454

FT CARBOHYD 1704 1704

SQ SEQUENCE 1743 AA; 194145 MW; 12DB363E2F729839 CRC64;

Query Match 29.0%; Score 54; DB 1; Length 265;

Best Local Similarity 45.5%; Pred. No. 16;

Matches 10; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 11 EIESNKKMLEKAYKESKNN 32

|||||

Db 116 ELDAKTKDLNSKTLQDNFKNN 137

|||||

RESULT 10

TAGC_DICDI STANDARD; PRT; 1743 AA.

AC Q23868;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Prestalk-specific protein tagc precursor (EC 3.4.21.-).

GN TAGC.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

OX NCBI_TaxID=44689;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AX4;

RX MEDLINE=97140317; PubMed=8986798;

RA Shaulsky G., Escalante R., Loomis W.F.;

RT "Developmental signal transduction pathways uncovered by genetic

RT suppressors";

RT Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).

CC -!- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGC MAY MEDIATE

INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS (BY

SIMILARITY).

-!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY

S8.

-!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING

TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.

-!- SIMILARITY: STRONG, TO TAGC.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; U60086; AAB03331.1;

DictyDB; D02060; tagc.

InterPro; IPR003593; AAA_AtpPase.

InterPro; IPR003439; ABC_transportr.

InterPro; IPR001140; ABCtransportr.

InterPro; IPR000209; Peptidase_S8.

Pfam; PF00005; ABC_tran; 1.

Pfam; PF00664; ABC_membrane; 1.

PRINTS; PR00723; SUBTILISIN.

SMART; SM00382; AAA; 1.

PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.

PROSITE; PS00137; SUBTILASE_HIS; 1.

PROSITE; PS00138; SUBTILASE_SER; FALSE_NEG.

PROSITE; PS00211; ABC_TRANSPORTER; 1.

KW Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;

KW Signal.

SIGNAL 1 27

CHAIN 28 1743

DOMAIN 316 642

DOMAIN ? 1743

TRANSMEM 962 982

TRANSMEM 1027 1047

TRANSMEM 1072 1092

TRANSMEM 1157 1177

TRANSMEM 1260 1280

TRANSMEM 1288 1308

ACT_SITE 325 325

ACT_SITE 372 372

ACT_SITE 637 637

NP_BIND 1495 1492

DOMAIN 42 46

DOMAIN 94 103

DOMAIN 643 646

DOMAIN 733 741

DOMAIN 786 792

DOMAIN 1337 1340

DOMAIN 1346 1352

DOMAIN 1353 1357

DOMAIN 1358 1364

DOMAIN 1381 1386

DOMAIN 1707 1729

CARBOHYD 390 390

CARBOHYD 536 536

CARBOHYD 547 547

CARBOHYD 614 614

CARBOHYD 689 689

CARBOHYD 735 735

CARBOHYD 741 741

CARBOHYD 776 776

CARBOHYD 832 832

CARBOHYD 887 887

CARBOHYD 1251 1251

CARBOHYD 1385 1385

CARBOHYD 1386 1386

CARBOHYD 1454 1454

CARBOHYD 1704 1704

SEQUENCE 1743 AA; 194145 MW; 12DB363E2F729839 CRC64;

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., R.E.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leivasalino M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McKurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.B., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.,
RT "The DNA sequence and comparative analysis of human chromosome 20";
RL Nature 414:865-871(2001).
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; N34668; AAA36528.1; -;
DR EMBL; X54130; CAA38065.1; -;
DR EMBL; X54890; CAA38662.1; -;
DR EMBL; X53364; CAA37447.1; -;
DR EMBL; AL121905; CAC10337.1; -;
DR PIR; A36065; A36065.
DR PIR; S12049; S12049.
DR HSP; P18052; IYFO.
DR Genew; HGNC:9664; PTPRA.
DR MIM; 176884; -;
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_pp.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE; PS00356; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Signal;
KW Repeat; Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 802
FT DOMAIN 20 142
FT TRANSMP 143 165
FT DOMAIN 166 500
FT DOMAIN 241 500
FT DOMAIN 501 802
FT ACT_SITE 442 442
FT ACT_SITE 732 732
FT CARBOHYD 21 21
FT CARBOHYD 36 36
FT CARBOHYD 68 68
FT CARBOHYD 80 80
FT CARBOHYD 86 86
FT CARBOHYD 104 104
FT CARBOHYD 124 124

FT VARSPLIC 139 147 MISSING (IN SHORT ISOFORM).
FT CONFLICT 114 114 T -> M (IN REF. 4).
FT CONFLICT 122 122 E -> P (IN REF. 5).
FT CONFLICT 138 138 S -> SGNSDKRR (IN REF. 2).
FT CONFLICT 179 187 MISSING (IN REF. 2, 3, 4, 5 AND 6).
FT CONFLICT 289 289 G -> E (IN REF. 4).
FT CONFLICT 367 367 V -> A (IN REF. 4).
FT CONFLICT 493 493 F -> S (IN REF. 4).
FT CONFLICT 786 786 K -> E (IN REF. 4).
SQ SEQUENCE 802 AA; 90599 MW; 8E964C3B56B5BE32 CRC64;
Query Match 28.8%; Score 53.5; DB 1; Length 802;
Best Local Similarity 35.3%; Pred. No. 61;
Matches 12; Conservative 8; Mismatches 11; Indels 3; Gaps 2;
Qy 3 FEPTPTNE-ETESNKKMLEKAYKESPKNGLP 35
Db 218 YPPLPVDKLEEEINRRMADNKKLFREF--NALP 249
RESULT 13
PTRA_MOUSE
ID PTRA_MOUSE STANDARD; PRT; 829 AA.
AC P18052; Q61808;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase alpha precursor (EC 3.1.3.48) (R-PTP-
DE alpha) (LCA-related phosphatase).
DE PTPRA OR LRP OR PTPA.
GN PTPRA OR LRP OR PTPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
RC STRAIN=C57BL/6 X DBA/2;
RX MEDLINE=90280391; PubMed=2162042;
RA Matthews R.J., Cahir E.D., Thomas M.L.;
RT "Identification of an additional member of the protein-tyrosine-
RT phosphatase family: evidence for alternative splicing in the tyrosine
RT phosphatase domain.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4444-4448(1990).
RN [2]
RP SEQUENCE OF 358-467 FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=95134232; PubMed=7832766;
RA Hendriks W., Schepens J., Brugman C., Zeeuwen P., Wieringa B.;
RT "A novel receptor-type protein tyrosine phosphatase with a single
RT catalytic domain is specifically expressed in mouse brain.";
RL Biochem. J. 305:499-504(1995).
RN [3]
RP SEQUENCE OF 651-756 FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=93086603; PubMed=1454056;
RA Schepens J., Zeeuwen P., Wieringa B., Hendriks W.;
RT "Identification and typing of members of the protein-tyrosine
RT phosphatase gene family expressed in mouse brain.";
RL Mol. Biol. Rep. 16:241-248(1992).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 202-503.
RX MEDLINE=96320562; PubMed=8700232;
RA Bilwes A.M., den Hertog J., Hunter T., Noel J.P.;
RT "Structural basis for inhibition of receptor protein-tyrosine
RT phosphatase-alpha by dimerization.";
RL Nature 382:555-559(1996).
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND
CC A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.


```
DE  IgA receptor precursor.
GN  ARP4.
OS  Streptococcus pyogenes.
OC  Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC  Streptococcus.
OX  NCBI_taxID=1314;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=AW43;
RX  MEDLINE=90113892; PubMed=2691841;
RA  Frithz E., Heden L.-O., Lindahl G.;
RT  "Extensive sequence homology between IgA receptor and M proteins in
RT  Streptococcus pyogenes.";
RL  Mol. Microbiol. 3:1111-1119(1989).
CC  -!- FUNCTION: BINDS IGA OF BOTH SUBCLASSES, AND ALSO BINDS POLYCLONAL
CC  IGG WEAKLY.
CC  -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan
CC  by an amide bond (Potential).
CC  -!- SIMILARITY: TO THE M PROTEINS OF STREPTOCOCCUS PYOGENES.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X15198; CAA33269.1; -
DR  PIR; S05568; S05568.
DR  InterPro; IPR001899; Gram_pos_anchor.
DR  InterPro; IPR003345; M_repeat.
DR  Pfam; PF00746; Gram_pos_anchor; 1.
DR  Pfam; PF02370; M; 7.
DR  PRINTS; PR00015; GP0SANCHOR.
DR  TIGRFAMS; TIGR01167; LPXTG_anchor; 1.
DR  TIGRFAMS; TIGR01168; YSIRK_signal; 1.
DR  PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
KW  Cell wall; Peptidoglycan-anchor; Receptor; Repeat; Signal.
FT  SIGNAL 1 41
FT  CHAIN 42 356 IGA RECEPTOR.
FT  PROPEP 357 386 REMOVED BY SORTASE (POTENTIAL).
FT  DOMAIN 42 152 IGA-BINDING (POTENTIAL).
FT  DOMAIN 111 158 GLN/GLU-RICH.
FT  DOMAIN 153 271 3 X TANDEM REPEATS.
FT  REPEAT 153 187 C-1.
FT  REPEAT 188 229 C-2.
FT  REPEAT 230 271 C-3.
FT  DOMAIN 327 352 GLY/PRO-RICH.
FT  SITE 353 357 LPXTG SORTING SIGNAL (POTENTIAL).
FT  MOD_RES 356 356 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ  SEQUENCE 386 AA; 43837 MW; C5FBDCCBB97BA088 CRC64;
```

```
Query Match      28.5%; Score 53; DB 1; Length 386;
Best Local Similarity 62.5%; Pred. No. 32;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
OY  9 NEEIESNKKMLEKEKA 24
    |:|:|:|:|:|:|
DB  283 NKEEGGKRLSEKKA 298
```

Search completed: April 29, 2003, 14:27:32
Job time : 29 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 14:22:45 ; Search time 29 Seconds
(without alignments)
248.678 Million cell updates/sec

Title: US-09-743-364-1

Perfect score: 186

Sequence: 1 FTFFPPTNEETESNKKMLEKAYKESFKNNGLP 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 205047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertibrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	181	97.3	149	16 Q99S08	Q99S08 staphylococ
2	58	31.2	192	5 Q9VNV9	Q9VNV9 drosophila
3	58	31.2	261	5 O15789	O15789 plasmodium
4	58	31.2	261	5 O18662	O18662 plasmodium
5	58	31.2	1650	5 O77328	O77328 plasmodium
6	57	30.6	431	5 Q21891	Q21891 caenorhabdi
7	57	30.6	445	5 Q9U389	Q9U389 caenorhabdi
8	57	30.6	522	2 Q9JMX7	Q9JMX7 helicobacte
9	57	30.6	522	16 O25263	O25263 helicobacte
10	57	30.6	522	16 Q9ZL09	Q9ZL09 helicobacte
11	57	30.6	731	4 O60433	O60433 homo sapien
12	56.5	30.4	781	10 O80743	O80743 arabidopsis
13	56.5	30.4	899	10 O8VY15	O8VY15 arabidopsis
14	56.5	30.4	1441	3 O74733	O74733 schizosacch
15	56	30.1	591	10 Q8VZ60	Q8VZ60 arabidopsis
16	56	30.1	695	10 Q9LU29	Q9LU29 arabidopsis

17	56	30.1	723	11 Q8VHF7	Q8VHF7 mus musculu
18	56	30.1	1400	11 Q9ESU6	Q9ESU6 mus musculu
19	56	30.1	1400	11 Q8VHF8	Q8VHF8 mus musculu
20	55	29.8	634	12 Q91LA5	Q91LA5 white spot
21	55	29.6	100	2 Q9R3A1	Q9R3A1 streptococc
22	55	29.6	198	2 Q54832	Q54832 streptococc
23	55	29.6	208	2 P95824	P95824 streptococc
24	55	29.6	208	2 P95825	P95825 streptococc
25	55	29.6	209	5 Q9U5G8	Q9U5G8 tetrahymena
26	55	29.6	288	12 Q91I73	Q91I73 turnip mosa
27	55	29.6	347	2 Q93RQ6	Q93RQ6 streptococc
28	55	29.6	408	2 Q54837	Q54837 streptococc
29	55	29.6	435	2 Q9AMM3	Q9AMM3 streptococc
30	55	29.6	436	2 Q54839	Q54839 streptococc
31	55	29.6	441	2 Q55246	Q55246 streptococc
32	55	29.6	443	2 Q54703	Q54703 streptococc
33	55	29.6	454	2 Q55278	Q55278 streptococc
34	55	29.6	457	2 Q54510	Q54510 streptococc
35	55	29.6	472	2 Q55279	Q55279 streptococc
36	55	29.6	475	2 Q33631	Q33631 streptococc
37	55	29.6	484	2 Q10372	Q10372 streptococc
38	55	29.6	484	16 Q99XV0	Q99XV0 streptococc
39	55	29.6	488	2 Q54830	Q54830 streptococc
40	55	29.6	500	2 Q9RHV2	Q9RHV2 streptococc
41	55	29.6	532	2 Q55098	Q55098 streptococc
42	55	29.6	539	2 Q54719	Q54719 streptococc
43	55	29.6	550	2 Q54840	Q54840 streptococc
44	55	29.6	558	2 Q54718	Q54718 streptococc
45	55	29.6	581	2 Q54835	Q54835 streptococc

ALIGNMENTS

RESULT 1

Q99S08	PRELIMINARY:	PRT:	149 AA.
ID	Q99S08		
AC	Q99S08;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Hypothetical protein SAI1755.		
GN	SAI1755.		
OS	Staphylococcus aureus (strain N315).		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;		
OC	Staphylococcus		
OX	NCBI_TaxID=158879;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21311952; PubMed=11418146;		
RA	Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,		
RA	Cui L., Oguchi A., Aoki K.I., Nagai Y., Iian J., Ito T., Kanamori M.,		
RA	Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,		
RA	Takanashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,		
RA	Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,		
RA	Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,		
RA	Ogasawara N., Hayashi H., Hiramatsu K.;		
RT	"Whole genome sequencing of methicillin-resistant Staphylococcus		
RT	aureus".		
RL	Lancet 357:1225-1240(2001).		
DR	EMBL; AP003135; BAB43029.1; .		
KW	Hypothetical protein; Complete proteome.		
SQ	SEQUENCE 149 AA; 17072 MW; 54EDE97E9CF36A4 CRC64;		

Query Match 97.3%; Score 181; DB 16; Length 149;
Best Local Similarity 97.1%; Pred. No. 1.4e-13;
Matches 34; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FTFFPPTNEETESNKKMLEKAYKESFKNNGLP 35

|||||

Db 29 FTFFPPTNEETESNKKMLEKAYKESFKNNGLP 63

```

RESULT 2
Q9VNV9 PRELIMINARY; PRT; 192 AA.
AC Q9VNV9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE CG14573 protein.
GN CG14573.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anandides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Frannk C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.E., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler A., Gabrielian A.C., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
DR EMBL; AF003595; AAF51774.1; -.
DR FlyBase; FBgn0037125; CG14573.
SQ SEQUENCE 192 AA; 20818 MW; D0E0D96436BD590B CRC64;

```

```

Query Match 31.2%; Score 58; DB 5; Length 192;
Best Local Similarity 52.6%; Pred. No. 31;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 2 TFEPTNEEIESNKKMLE 20
DB 87 TFEPTNEEIVDEESLE 105

```

```

RESULT 3
O15789 PRELIMINARY; PRT; 261 AA.
ID O15789
AC O15789;

```

```

DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE CG9.
DE CG9.
OS Plasmodium falciparum (isolate 7G8).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57266;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7G8;
RX MEDLINE=98054002; PubMed=9393853;
RA Su X., Kirkman L.A., Fujioka H., Wellem T.E.;
RT "Complex polymorphisms in an approximately kDa protein are linked to
RT chloroquine-resistant P. falciparum in Southeast Asia and Africa."
RL Cell 91:593-603(1997).
DR EMBL; AF030691; AAC47849.1; -.
SQ SEQUENCE 261 AA; 31065 MW; 64F83824239EFFB3 CRC64;

```

```

Query Match 31.2%; Score 58; DB 5; Length 261;
Best Local Similarity 37.0%; Pred. No. 42;
Matches 10; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 6 FPTNEEIESNKKMLEKAYKESFKNN 32
DB 83 FDEHKEINNNNNKKKSYNNYKNN 109

```

```

RESULT 4
O18662 PRELIMINARY; PRT; 261 AA.
ID O18662
AC O18662;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CG9 protein.
DE CG9.
OS Plasmodium falciparum (isolate Dd2).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57267;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DD2, AND HB3;
RX MEDLINE=98054002; PubMed=9393853;
RA Su X.-Z., Kirkman L.A., Fujioka H., Wellem T.E.;
RT "Complex polymorphisms in an approximately kDa protein are linked to
RT chloroquine-resistant P. falciparum in Southeast Asia and Africa."
RL Cell 91:593-603(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DD2, AND HB3;
RA Su X.-Z., Kirkman L.A., Wellem T.E.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF030694; AAC47839.1; -.
DR EMBL; AF030690; AAC47852.1; -.
SQ SEQUENCE 261 AA; 31108 MW; 40B69895171CE4BC CRC64;

```

```

Query Match 31.2%; Score 58; DB 5; Length 261;
Best Local Similarity 37.0%; Pred. No. 42;
Matches 10; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 6 FPTNEEIESNKKMLEKAYKESFKNN 32
DB 83 FDEHKEINNNNNKKKSYNNYKNN 109

```

```

RESULT 5
O77328 PRELIMINARY; PRT; 1650 AA.
ID O77328
AC O77328;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

```

```

QY      8  TNEETESNKKMLEKAYKESFK 30
      : | | | | | : | | :
DB      216 SNHEITENKKLSEKMEVKNKEFE 238

RESULT 7
ID Q90389 PRELIMINARY; PRT; 445 AA.
AC Q90389;
DT 01-NAY-2000 (TremBLRel. 13, Created)
DT 01-NAY-2000 (TremBLRel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLRel. 17, Last annotation update)
DE R102.5b protein.
DE R102.5B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RL Berks M.;
RA Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RT none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RT Science 282:2012-2018(1998).
DR EMBL; 270309; CAB54290.1; -.
SQ SEQUENCE 445 AA; 50381 MW; 5AC5FDB8D65BA61B CRC64;

Query Match 30.6%; Score 57; DB 5; Length 445;
Best Local Similarity 47.8%; Pred. No. 91;
Matches 11; Conservative 4; Mismatches 8; Indels 0; Gaps 0

QY      8  TNEETESNKKMLEKAYKESFK 30
      : | | | | | : | | :
DB      216 SNHEITENKKLSEKMEVKNKEFE 238

RESULT 8
ID Q9JMX7 PRELIMINARY; PRT; 522 AA.
AC Q9JMX7;
DT 01-OCT-2000 (TremBLRel. 15, Created)
DT 01-OCT-2000 (TremBLRel. 15, Last sequence update)
DT 01-JUN-2002 (TremBLRel. 21, Last annotation update)
DE Cag-X.
DE CAG-X.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CCUG 17874;
RX MEDLINE=97121442; PubMed=8962108;
RA Censini S., Lange C., Xiang Z., Crabtree J.E., Ghiara P.,
RA Borodovsky M., Rappuoli R., Covacci A.;
RT "cag, a pathogenicity island of Helicobacter pylori, encodes type I-
RT specific and disease-associated virulence factors.";
RT Proc. Natl. Acad. Sci. U.S.A. 93:14648-14653(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX 'STRAIN=CCUG 17874;
RX MEDLINE=20150112; PubMed=10684850;
RA Covacci A., Rappuoli R.;
RT "Tyrosine-phosphorylated bacterial proteins: Trojan horses for the
RT host cell.";
RT J. Exp. Med. 191:587-592(2000).
DR EMBL; AF282852; AAF0199.1; -.
DR InterPro; IPR004357; IYSec_CXag

```

```

DR Pfam; PF03524; cagX; 1.
DR PRINTS; PR01556; TYPE4SSCAGX.
SQ SEQUENCE 522 AA; 60671 MW; 0B0F8E796339B4E2 CRC64;

Query Match
  30.6%; Score 57; DB 2; Length 522;
Best Local Similarity 57.1%; Pred. No. 1.1e+02;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 7 PTNEEIESNKKMLEKEKAYKE 27
  | : : | | | | | | | |
Db 136 PDPKELEEQKKALEKEKEAKE 156

RESULT 9
O25263
ID O25263 PRELIMINARY; PRT; 522 AA.
AC O25263;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CAG pathogenicity island protein (CAG8).
GN HP0528.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=97394467; PubMed=9252185;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Ufford S.D., deJonge B.L., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL; AE001481; AAD06048.1; -.
DR InterPro; IPR004357; IVSec_cagX.
DR Pfam; PF03524; cagX; 1.
DR PRINTS; PR01556; TYPE4SSCAGX.
KW Complete proteome.
SQ SEQUENCE 522 AA; 60593 MW; A9A48597CFD6C882 CRC64;

Query Match
  30.6%; Score 57; DB 16; Length 522;
Best Local Similarity 57.1%; Pred. No. 1.1e+02;
Matches 12; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 7 PTNEEIESNKKMLEKEKAYKE 27
  | : : | | | | | | | |
Db 136 PDPKELEEQKKALEKEKEAKE 156

RESULT 11
O60433
ID O60433 PRELIMINARY; PRT; 731 AA.
AC O60433;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE R31546_1 (Fragment).
GN HUNK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RX Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K.J., Gordon L., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Barnes J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Frankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of an ~1.5 Mb OLFR cluster in 19p13.1.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004798; AAC27978.1; -.
DR HSSP; Q92831; 1B91.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
FT NON_TER
SQ SEQUENCE 731 AA; 81383 MW; F56C3A02F1A26F65 CRC64;

Query Match
  30.6%; Score 57; DB 4; Length 731;
Best Local Similarity 44.8%; Pred. No. 1.5e+02;
Matches 13; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 4 EPFPTNEEIESNKKMLEKEKAYKESFKN 32
  | : : | | | | | | | |
Db 560 EKHKRKEVEENKSKAKEPPPKTKKN 588

RESULT 12
O80743

```

```

DR InterPro; IPR002483; PWI.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF01480; PWI; 1.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00311; PWI; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
KW Hypothetical protein.
SQ SEQUENCE 899 AA; 101497 MW; 2C9035D065C0E89B CRC64;
Query Match 30.4%; Score 56.5; DB 10; Length 899;
Best Local Similarity 44.8%; Pred.No.2e+02;
Matches 13; Conservative 8; Mismatches 7; Indels 1
QY 8 TNEIESNKKMLEK-EKAYKEFSKNNGLP 35
Db 349 TDEREADREAMEKETETAERLKSNNLP 377
|:| |::: |:| |:| |:| |:| |:|
RESULT 14
ID 074733 PRELIMINARY; PRT: 1441 AA.
AC 074733;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 160.3 kDa protein C1709.08 in chromosome II.
GN SPBC1709.08.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCBI_TaxID=4896;
RX [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL031852; CAA21247.1;
DR InterPro; IPR004871; CPSF_A.
DR Pfam; PF03178; CPSF_A; 1.
DR KW Hypothetical protein; Transmembrane.
FT TRANSMEM 561 581 POTENTIAL.
SQ SEQUENCE 1441 AA; 160299 MW; C89E516A10C686B CRC64;
Query Match 30.4%; Score 56.5; DB 3; Length 1441;
Best Local Similarity 36.7%; Pred.No.3.2e+02;
Matches 11; Conservative 6; Mismatches 8; Indels 5
QY 5 PPTNEEIESNKKMLEKEK-----AYKESF 29
Db 189 PYPANEDLDMEEAATENSIISSSYAYKPSP 218
|:| |::: |:| |:| |:| |:| |:|
RESULT 15
Q8VZ60
ID Q8VZ60 PRELIMINARY; PRT: 591 AA.
AC Q8VZ60;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative translation initiation factor eIF3 protein.
GN A74G20980.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheopsida;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
ON NCBI_TaxID=3702;
RX [1]
RN RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith
RA Lee J.-M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H.,

```

RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narasaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shin P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RA "Full Length cDNA of gene At4g20980 (GI:152333469).";
RT Submitted (DEC-2001) to the At4g20980/DBJ databases.
RL EMBL; AY05228; AAL38704.1; -;
RW Initiation factor.
SQ SEQUENCE 591 AA; 66738 MW; A833D9F6334B084D CRC64;
KW

Query Match 30.1%; Score 56; DB 10; Length 591;
Best Local Similarity 39.3%; Pred. No. 1.6e+02;
Matches 11; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Qy 5 PPTNEEIESNKKMLEKEKAYKESFKNN 32
- :||:| : ||:| : ||
Db 139 PORRDEEVEAKKRDAEKERARRDRLYNN 166

Search completed: April 29, 2003, 14:28:08
Job time : 32 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	55	29.6	236	3	US-08-937-271-11	Sequence 11, Appl
2	55	28.6	254	4	US-08-914-479A-6	Sequence 4, Appl
3	55	29.6	284	4	US-08-914-479A-6	Sequence 6, Appl
4	55	29.6	305	3	US-08-937-271-10	Sequence 10, Appl
5	55	29.6	443	2	US-08-795-475-6	Sequence 6, Appl
6	53.5	28.8	793	1	US-08-015-985-3	Sequence 3, Appl
7	53.5	28.8	802	1	US-08-015-985-1	Sequence 1, Appl
8	53	28.5	344	6	5210183-2	Patent No. 5210183
9	53	28.5	683	6	5210183-3	Patent No. 5210183
10	52.5	28.2	715	3	US-08-425-843-7	Sequence 7, Appl
11	50.5	27.2	402	4	US-09-134-001C-4674	Sequence 4674, Ap
12	50	26.9	318	3	US-09-188-579-81	Sequence 81, Appl
13	50	26.9	318	4	US-09-315-444-81	Sequence 81, Appl
14	50	26.9	318	4	US-09-721-362-81	Sequence 81, Appl
15	49	26.3	246	1	US-07-887-072B-4	Sequence 4, Appl
16	49	26.3	246	1	US-08-466-444-4	Sequence 4, Appl
17	49	26.3	740	1	US-08-257-073-5	Sequence 5, Appl
18	48.5	26.1	332	2	US-08-405-175A-5	Sequence 5, Appl
19	48.5	26.1	335	2	US-08-405-175A-6	Sequence 5, Appl
20	48.5	26.1	427	4	US-09-134-001C-5143	Sequence 5143, Ap
21	48	25.8	210	4	US-09-222-938A-67	Sequence 67, Appl
22	48	25.8	591	2	US-08-736-770-5	Sequence 5, Appl
23	48	25.8	608	2	US-08-736-770-1	Sequence 1, Appl
24	47	25.3	366	3	US-08-860-368B-20	Sequence 20, Appl
25	47	25.3	1066	2	US-08-308-818-1	Sequence 1, Appl
26	47	25.3	2522	4	US-09-251-645-13	Sequence 13, Appl
27	46.5	25.0	170	4	US-09-129-030-40	Sequence 40, Appl

```
RESULT 2
US-08-914-479A-4
; Sequence 4, Application US/08914479A
; Patent No. 6419932
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 481112.404C2
; CURRENT APPLICATION NUMBER: US/08/914,479A
; CURRENT FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 08/409,270
; PRIOR FILING DATE: 1995-03-23
; PRIOR APPLICATION NUMBER: 07/945,860
; PRIOR FILING DATE: 1992-09-16
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: An antigen of M5 and a carrier of the
; OTHER INFORMATION: COOH-terminal portion of M5
US-08-914-479A-4
Query Match          29.6%; Score 55; DB 4; Length 254;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 NEEIESNKKMLEKEKA 24
   1:11:111:111111
Db 131 NKELESKKLTKTEKA 146

RESULT 3
US-08-914-479A-6
; Sequence 6, Application US/08914479A
; Patent No. 6419932
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 481112.404C2
; CURRENT APPLICATION NUMBER: US/08/914,479A
; CURRENT FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 08/409,270
; PRIOR FILING DATE: 1995-03-23
; PRIOR APPLICATION NUMBER: 07/945,860
; PRIOR FILING DATE: 1992-09-16
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: An antigen of three fragments of M5 and a carrier
; OTHER INFORMATION: of the COOH-terminal portion of M5
US-08-914-479A-6
Query Match          29.6%; Score 55; DB 4; Length 284;
Best Local Similarity 62.5%; Pred. No. 13;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 NEEIESNKKMLEKEKA 24
   1:11:111:111111
Db 161 NKELESKKLTKTEKA 176

RESULT 4
US-08-937-271-10
; Sequence 10, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-10
Query Match          29.6%; Score 55; DB 3; Length 305;
Best Local Similarity 62.5%; Pred. No. 14;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 NEEIESNKKMLEKEKA 24
   1:11:111:111111
Db 182 NKELESKKLTKTEKA 197

RESULT 5
US-08-795-475-6
; Sequence 6, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvrck, Lars
; APPLICANT: Sjdbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```

; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-475-6
Query Match 29.6%; Score 55; DB 2; Length 443;
Best Local Similarity 62.5%; Pred. No. 20;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 NEEIESNKKMLEKEKA 24
Db 320 NKELESKKLKEKEKA 335

RESULT 6
US-08-015-985-3
; Sequence 3, Application US/08015985
; Patent No. 5538886
; GENERAL INFORMATION:
; APPLICANT: Schlensing, Joseph
; APPLICANT: Sap, Jan M.
; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-ALPHA
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/015,985
; FILING DATE: 10-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-015-985-1
Query Match 28.8%; Score 53.5; DB 1; Length 802;
Best Local Similarity 35.3%; Pred. No. 60;
Matches 12; Conservative 8; Mismatches 11; Indels 3; Gaps 2;

Qy 3 FEPPPTNE-EIESNKKMLEKEKAYKESFKNNGLP 35
Db 218 YPPLPVDKLEIEINRRMADNKLFREEF--NALP 249

RESULT 8
5210183-2
; Patent No. 5210183
; APPLICANT: LINDAHL, GUNNAR;FRITHZ, ELISABET;HEDEN, LARS-OLOF
; TITLE OF INVENTION: PROTEIN ARP, WITH IMMUNOGLOBULIN A
; BINDING ACTIVITY. THE CORRESPONDING VECTORS AND HOSTS, REAGENT
; KIT AND PHARMACEUTICAL COMPOSITION
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/270,099
; FILING DATE: 14-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 186,097
; FILING DATE: 25-APR-1988
; SEQ ID NO:2:
; LENGTH: 344
5210183-2
Query Match 28.5%; Score 53; DB 6; Length 344;
Best Local Similarity 62.5%; Pred. No. 28;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```
QY 9 NEEIESNKKMLEKEKA 24
      1:1:1 11:11111
Db 241 NKELEGKLLSEKEKA 256

RESULT 9
5210183-3
; PATENT NO. 5210183
; APPLICANT: LINDAHL, GUNNAR;FRITZ, ELISABET;HEDEN, LARS-OLOF
; TITLE OF INVENTION: PROTEIN ARP, WITH IMMUNOGLOBULIN A
; BINDING ACTIVITY, THE CORRESPONDING VECTORS AND HOSTS, REAGENT
; KIT AND PHARMACEUTICAL COMPOSITION
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/270,099
; FILING DATE: 14-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 186,097
; FILING DATE: 25-APR-1988
; SEQ ID NO:3:
; LENGTH: 683
5210183-3

Query Match      28.5%; Score 53; DB 6; Length 683;
Best Local Similarity 62.5%; Pred. No. 59;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 9 NEEIESNKKMLEKEKA 24
      1:1:1 11:11111
Db 583 NKELEGKLLSEKEKA 598

RESULT 10
US-08-425-843-7
; Sequence 7, Application US/08425843
; Patent No. 6020154
; GENERAL INFORMATION:
; APPLICANT: Hansen, Eric J.
; APPLICANT: Cope, Leslie D.
; APPLICANT: Jarosik, Gregory P.
; APPLICANT: Hanson, Mark S.
; TITLE OF INVENTION: H. Influenzae Hxub and HxuC Genes, Proteins
; TITLE OF INVENTION: and Methods of Use
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,843
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: AMCY:012/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 715 amino acids
; TYPE: amino acid

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-425-843-7

Query Match      28.2%; Score 52.5; DB 3; Length 715;
Best Local Similarity 37.5%; Pred. No. 71;
Matches 12; Conservative 9; Mismatches 6; Indels 5; Gaps 2;

QY 7 PTNEEIE---SNKKMLEKEKAYKESFKNNGLP 35
      1:1:1 11:11111
Db 247 PSNNEVENELTKIIDIQINEFHGS--NNGLP 276

RESULT 11
US-09-134-001C-4674
; Sequence 4674, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4674
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4674

Query Match      27.2%; Score 50.5; DB 4; Length 402;
Best Local Similarity 45.2%; Pred. No. 68;
Matches 14; Conservative 5; Mismatches 9; Indels 3; Gaps 1;

QY 4 EPFTNEEIESNKKMLEKEKAYKESFKNNGL 34
      1:1:1 11:11111
Db 120 DPYTKREID---KMSEKKEKSKNNANENLGL 147

RESULT 12
US-09-188-579-81
; Sequence 81, Application US/09188579B
; Patent No. 6107040
; GENERAL INFORMATION:
; APPLICANT: Shuman, Stewart
; TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
; FILE REFERENCE: D6185
; CURRENT APPLICATION NUMBER: US/09/188,579B
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 114
; SEQ ID NO 81
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of RNA guanylyltransferase.
US-09-188-579-81

Query Match      26.9%; Score 50; DB 3; Length 318;
Best Local Similarity 38.7%; Pred. No. 61;
Matches 12; Conservative 8; Mismatches 9; Indels 2; Gaps 1;

QY 6 FPTNEEIESNKKMLE--KEKAYKESFKNNGL 34
      1:1:1 11:11111
Db 1 FPGSQEVSFSSKKHLQALKERNYFVCEKSDGI 31
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 14:28:10 ; Search time 16 Seconds
(without alignments)
175.284 Million cell updates/sec

Title: US-09-743-364-1

Perfect score: 186

Sequence: 1 FTFFPPTNEIESNKKMLEKEKAYKESFKNGLP 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB_PEP.p:
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB_PEP.p:
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB_PEP.p:
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB_PEP.p:
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB_PEP.p:
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB_PEP.p:
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB_PEP.p:
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	55	29.6	254	9	US-10-141-627-4
2	55	29.6	284	9	US-10-141-627-6
3	55	29.6	443	1	US-08-325-278-6
4	55	29.6	615	9	US-09-738-626-6067
5	55	29.6	1080	10	US-09-904-380-2
6	52	28.0	544	9	US-09-925-299-1004
7	52	28.0	544	10	US-09-925-299-1004
8	52	28.0	1464	9	US-09-842-777-10
9	51.5	27.7	1400	10	US-09-764-176-7
10	50.5	27.2	278	10	US-09-767-041-36
11	50.5	27.2	332	10	US-09-767-041-22
12	50.5	27.2	640	9	US-10-243-735-2
13	50.5	27.2	654	10	US-09-940-921B-4
14	50.5	27.2	683	10	US-09-940-921B-2
15	50.5	27.2	1163	9	US-09-932-257A-3
16	50	26.9	354	10	US-09-815-242-5345
17	50	26.9	444	10	US-09-815-242-12483
18	50	26.9	484	9	US-09-738-626-5243
19	49.5	26.6	60	12	US-10-001-843-191

20 49 26.3 76 9 US-10-066-127-6 Sequence 6, Appli
21 49 26.3 194 9 US-10-025-380-1126 Sequence 1136, Ap
22 48.5 26.1 374 10 US-09-925-302-711 Sequence 711, App
23 48.5 26.1 1323 10 US-09-801-368-34 Sequence 34, Appl
24 48 25.8 54 9 US-09-820-843A-110 Sequence 110, Appl
25 48 25.8 232 10 US-09-803-286A-6 Sequence 6, Appli
26 48 25.8 2665 10 US-09-864-761-34248 Sequence 34248, A
27 47 25.3 233 9 US-09-925-299-1014 Sequence 1014, Ap
28 47 25.3 231 10 US-09-925-299-1014 Sequence 1014, Ap
29 47 25.3 241 10 US-09-925-301-1105 Sequence 1105, Ap
30 47 25.3 494 10 US-09-764-864-1235 Sequence 1235, Ap
31 47 25.3 2516 10 US-09-817-514A-2 Sequence 2, Appli
32 46.5 25.0 700 9 US-10-153-273-10 Sequence 10, Appl
33 46.5 25.0 736 10 US-09-815-242-12646 Sequence 12646, A
34 46 24.7 71 10 US-09-864-761-38143 Sequence 38143, A
35 46 24.7 191 10 US-09-733-507-2 Sequence 2, Appli
36 46 24.7 191 10 US-09-733-507-10 Sequence 10, Appl
37 46 24.7 239 10 US-09-815-242-5352 Sequence 5352, Ap
38 46 24.7 253 10 US-09-815-242-12278 Sequence 12278, A
39 46 24.7 308 9 US-10-107-461-2 Sequence 2, Appli
40 46 24.7 308 10 US-09-816-664-5 Sequence 5, Appli
41 46 24.7 555 10 US-09-764-864-1161 Sequence 1161, Ap
42 46 24.7 555 10 US-09-764-864-1574 Sequence 1574, Ap
43 46 24.7 574 10 US-09-391-340-12 Sequence 12, Appl
44 46 24.7 574 10 US-09-948-369-12 Sequence 12, Appl
45 46 24.7 1192 10 US-09-815-242-10903 Sequence 10903, A

ALIGNMENTS

RESULT 1
US-10-141-627-4
; Sequence 4, Application US/10141627
; Patent No. US20020176863A1
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 481112.404C3
; CURRENT APPLICATION NUMBER: US/10/141,627
; CURRENT FILING DATE: 2002-05-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 254
; TYPE: PRT
; FEATURE:
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: An antigen of M5 and a carrier of the
; OTHER INFORMATION: COOH-terminal portion of M5
US-10-141-627-4

Query Match 29.6%; Score 55; DB 9; Length 254;
Best Local Similarity 62.5%; Pred. No. 17;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 NEEIESNKKMLEKEKA 24

Db 131 NKEEESKKLTEREKA 146

RESULT 2

US-10-141-627-6
; Sequence 6, Application US/10141627
; Patent No. US20020176863A1
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 481112.404C3
; CURRENT APPLICATION NUMBER: US/10/141,627
; CURRENT FILING DATE: 2002-05-07

```
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: An antigen of three fragments of M5 and a carrier
; OTHER INFORMATION: of the COOH-terminal portion of M5
US-10-141-627-6

Query Match          29.6%; Score 55; DB 9; Length 284;
Best Local Similarity 62.5%; Pred. No. 19;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 NEEIESNKKMLEKEKA 24
Db 161 NKELESKKLTKLEKEKA 176

RESULT 3
US-08-325-278-6
; Sequence 6, Application US/08325278
; Publication No. US20030027283A1
; GENERAL INFORMATION:
; APPLICANT: Bjvick, Lars
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,278
; FILING DATE: 26-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 450023.401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-278-6

Query Match          29.6%; Score 55; DB 1; Length 443;
Best Local Similarity 62.5%; Pred. No. 32;
Matches 10; Conservative 4; Mismatches 2; Indels 2; Gaps 0;

Qy 9 NEEIESNKKMLEKEKA 24
Db 320 NKELESKKLTKLEKEKA 335

RESULT 4
US-09-738-626-6067
; Sequence 6067, Application US/09738626
; Publication No. US20020197609A1
```

```
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKOHO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6067
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6067

Query Match          29.6%; Score 55; DB 9; Length 615;
Best Local Similarity 47.8%; Pred. No. 47;
Matches 11; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 11 EIESNKKMLEKEKAYKESFKNNG 33
Db 575 DISRKKLLEKQKAGKKRKNKNG 597

RESULT 5
US-09-904-380-2
; Sequence 2, Application US/09904380
; Patent No. US20020022229A1
; GENERAL INFORMATION:
; APPLICANT: Jane H. Morse and James A. Knowles
; TITLE OF INVENTION: Role of pp11 Gene in Pulmonary Hypertension
; FILE REFERENCE: 0575/62430-A/JPW/SHS
; CURRENT APPLICATION NUMBER: US/09/904,380
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1080
; TYPE: PRT
; ORGANISM: Human
; MOLECULE TYPE: protein
US-09-904-380-2

Query Match          29.6%; Score 55; DB 10; Length 1080;
Best Local Similarity 48.3%; Pred. No. 89;
Matches 14; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

Qy 8 TNEEIESNKKMLEKE--KAYKESFKNNG 34
Db 684 TEEDLETNKKLDPKVDKNLKESSDENNL 712

RESULT 6
US-09-925-299-1004
; Sequence 1004, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
```


RESULT 8
US-09-842-777-10
; Sequence 10, Application US/09842777
; Publication No. US2002018268A1
; GENERAL INFORMATION:

```

RESULT 10
US-09-767-041-36
; Sequence 36, Application US/09767041
; Patent No. US2002005518A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22

```

```

; CURRENT APPLICATION NUMBER: US/09/940-921B
; CURRENT FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: US 60/229,280
; PRIOR FILING DATE: 2000-08-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 683
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-940-921B-2

Query Match          27.2%; Score 50.5; DB 10; Length 683;
Best Local Similarity 39.5%; Pred. No. 1.9e+02;
Matches 15; Conservative 5; Mismatches 11; Indels 7; Gaps 2;

QY      3  FEPPPTNEETESNKK--MLEKEK-----AYKESFKNG 33
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

DB 37 FEKMPIQEKASKKEVILLEKKMKHPNIVAFNFSFOENG 74

RESULT 15
US-09-932-257A-3
; Sequence 3, Application US/09932257A
; Publication No. US20030039658A1
; GENERAL INFORMATION:
; APPLICANT: Estable, Mario
; APPLICANT: Roeder, Robert
; TITLE OF INVENTION: MCEF, A No. US20030039658A1el Transcription Factor
; FILE REFERENCE: 600-1-269N
; CURRENT APPLICATION NUMBER: US/09/932,257A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/226,340
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 60/226,339
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1086)..(1086)
; OTHER INFORMATION: x is any amino acid
US-09-932-257A-3

Query Match 27.2%; Score 50.5; DB 9; Length 1163;
Best Local Similarity 44.8%; Pred. No. 3.5e+02;
Matches 13; Conservative 4; Mismatches 11; Indels 1; Gaps 1;
QY 5 PPPTNEETESNKKMLEKKAYKESFKNG 33
|||||.:||.:|.:|.:|.:|
DB 346 PPPTKESQSNFGE-QRYNPSTSG 373

Search completed: April 29, 2003, 14:36:11
Job time : 17 secs

THIS PAGE BLANK (USPTO)